

Saturday, October 25, 10:00 AM - 11:15 AM

Building A Level 3 A315

An Introduction to Modeling with GAMS Py

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - An Introduction to Modeling with GAMS Py

Steven Dirkse, GAMS Development Corporation, Fairfax, VA, United States, Adam Christensen

GAMS execution system with the flexible Python language, creating a powerful mathematical optimization package. It acts as a bridge between the expressive Python language and the robust GAMS system, allowing you to effortlessly create complex mathematical models and applications. Join us to explore GAMS Py's fundamental functionalities through practical, interactive exercises. We'll cover everything from defining sets, parameters, variables, and equations to solving models and retrieving results, all within a familiar Python environment. Beyond the basics, we'll also provide a glimpse into more advanced features, demonstrating how GAMS Py can streamline complex modeling workflows and enhance your analytical capabilities. Whether you're a seasoned GAMS user looking to integrate with Python or a Python user curious about optimization, this workshop will equip you with essential skills needed to get started and demonstrate what is possible with GAMS Py.

Saturday, October 25, 11:30 AM - 12:45 PM

Building A Level 3 A315

The Benefits of Certifying your Data Team

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - The Benefits of Certifying Your Data Team

Nicole Janeway Bills, Data Strategy Professionals, Alpharetta, GA, United States

Certification elevates a data team's capability, credibility, and business impact. Data Strategy Professionals will explore how recognized credentials in Data Strategy, Data Governance, Data Analytics, and related disciplines strengthen technical skills, improve cross-departmental trust, and support regulatory compliance. Participants will learn how certification fosters a shared vocabulary, boosts productivity, and aligns data work with strategic objectives—resulting in higher quality outcomes and measurable business value.

Building A Level 3 A316

SAS Analytics

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - SAS Analytics

Rob Pratt, SAS, Cary, NC, United States

SAS offers extensive analytic capabilities, including machine learning, deep learning, natural language processing, statistical analysis, optimization, and simulation. SAS analytic functionality is also available through the open, cloud-enabled design of SAS® Viya®. You can program in SAS or in other languages – Python, Lua, Java, and R. This workshop will review the SAS analytics portfolio, highlight recently added optimization features, and explore several optimization examples. Academic faculty, staff, and students can access the latest SAS software through our free academic software platforms SAS Viya for Learners and SAS Viya Workbench for Learners. Access instructions and learning resources will be provided.

Saturday, October 25, 1:00 PM - 2:15 PM

Building A Level 3 A315

How to Advance your Optimization Game with FICO® Xpress Solver

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - How to Advance Your Optimization Game with FICO(R) Xpress Solver

Tristan Gally, FICO, Xpress Optimization, Birmingham, United Kingdom, Michael Perregaard, Susanne Heipcke, Dinakar Gade

In this workshop, we will give an overview of the different solvers in the FICO Xpress Optimization suite, their features, and how to interact with them through the different APIs. Find out how Xpress can solve LPs, MIPs, NLPs, and general MINLPs to global optimality, get an overview of different cuts, heuristics, and other features and learn how to influence the solve using callbacks. We will also give an overview of recent advances to all the different solvers from MIP over hybrid gradient to the FICO Xpress Global solver.

Building A Level 3 A316

5 Pillars of OR Project Success: A DecisionOps Tutorial with Nextmv

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - 5 Pillars of OR Project Success: a DecisionOPS Tutorial with Nextmv

Tiffany Bogich, Nextmv, Salt Lake City, UT, United States, Carolyn Mooney

Why do OR projects fail? Why do they succeed? The way a project unfolds often comes down to how a project is run. DecisionOps (yes, it's like MLOps, but for decision science) makes all the difference because: People don't trust decisions they can't understand. Every optimization run needs to be logged, tracked, and explainable Testing reduces rollout risk and increases revenue opportunities Structured workflows across a diverse, integrated ecosystem provides choice and resilience These four things amount to: observability, explainability, systematic testing, and consistency in support of diversity. And they all depend upon a robust platform that facilitates collaboration across modelers, engineers, and business users. In this interactive workshop, you'll learn how to efficiently operationalize open source and commercial decision models, create workflows that link to up and downstream processes like ML forecasting and result visualization, perform robust scenario testing to explore what

if questions, and manage model versions and run data. What's more: You have the opportunity to see all of this in action — and do it yourself live. This session is relevant to data science and OR practitioners using tooling such as OR-Tools, Pyomo, HiGHS, VROOM, Gurobi, AMPL, FICO Xpress, Hexaly, Statsmodels, scikit-learn, Databricks, Snowflake, and more.

Saturday, October 25, 2:30 PM - 3:45 PM

Building A Level 3 A316

AI + Monte Carlo Simulation: Transforming Risk Analysis with GenAI

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - AI + Monte Carlo Simulation: Transforming Risk Analysis with GenAI

Gustavo Vinueza, Ferryfield, Ithaca, NY, United States

Learn how to integrate Generative AI into Monte Carlo Simulation to build smarter, and more understandable risk models. This session will showcase how to enhance their risk analysis workflows by integrating the power of Generative AI (GenAI) into Monte Carlo Simulation (MCS). You'll gain practical skills and insights that bridge the gap between modern AI tools and traditional quantitative risk methods. Main cases to be analyzed: 1) Selecting the Right Distribution, 2) Use Case 2 – Creating Risk Registers, 3) Explaining a Histogram to Non-Experts, 4) Writing Executive Summaries.

Building A Level 3 A315

Part 1: Pushing the Boundaries of Optimization: What's New in Gurobi 13 and Beyond

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Part 1: Pushing the Boundaries of Optimization: What's New in Gurobi 13 and Beyond

Robert Luce, Gurobi, Beaverton, OR, United States, Ed Klotz, Dan Steffy

Join Gurobi's expert team for a deep dive into the latest advancements in mathematical optimization. This workshop will explore new features in Gurobi 13, including a CPU-based PDHG solver and a local nonlinear solver. Gain insights into emerging technologies like generative AI for optimization, quantum computing, and our progress on GPU acceleration. The session will conclude with a dynamic panel discussion on the future of optimization and AI. Whether you're an academic, practitioner, or student, this is your chance to engage directly with Gurobi's developers and expand your understanding of cutting-edge optimization tools.

Building A Level 3 A304

No-code Machine Learning with JMP Pro and JMP Student Edition

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - No-Code Machine Learning with JMP Pro and JMP Student Edition

Ross Metusalem, JMP Statistical Discovery, Tampa, FL, United States

Building and deploying machine learning models doesn't have to require expertise in Python or other coding languages. JMP Student Edition (free for academic use at jmp.com/student) and JMP Pro offer interactive, point-and-click tools for building powerful machine learning models for regression and classification, from basic decision trees and neural nets through XGBoost and deep learning models. Join this workshop to learn how to use these no-code tools for: Training machine learning models for regression and classification, including image classification Automated cross-validation and hyperparameter tuning. Model comparison and deployment.

Building A Level 3 A303

Timefold: The Open Source Solver for Java

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Timefold: The Open Source Solver for Java

Jeroen Colin, Timefold BV, Gent, Belgium, Geoffrey De Smet

Timefold Solver is a metaheuristics solver for vehicle routing, shift rostering, job scheduling and other planning problems. It's designed for ease of use, scalability and enterprise usage. But unlike other solvers, it doesn't require you to input your constraints/objectives as math equations. It accepts them as plain Java code. They can reuse existing code, for example from date and time libraries. And you don't face any linear or quadratic limitations. In this session, we'll introduce the open source solver, explain the basics and live code a solution for school scheduling.

Saturday, October 25, 4:00 PM - 5:15 PM

Building A Level 3 A315

Part 2: Pushing the Boundaries of Optimization: What's New in Gurobi 13 and Beyond

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Part 2: Pushing the Boundaries of Optimization: What's New in Gurobi 13 and Beyond

Dan Steffy, Gurobi Optimization, Bloomfield Hills, MI, United States, Ed Klotz, Robert Luce

Join Gurobi's expert team for a deep dive into the latest advancements in mathematical optimization. This workshop will explore new features in Gurobi 13, including a CPU-based PDHG solver and a local nonlinear solver. Gain insights into emerging technologies like generative AI for optimization, quantum computing, and our progress on GPU acceleration. The session will conclude with a dynamic panel discussion on the future of optimization and AI. Whether you're an academic, practitioner, or student, this is your chance to engage directly with Gurobi's developers and expand your understanding of cutting-edge optimization tools.

Building A Level 3 A304

MDPI Academic Publishing Workshop

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - MDPI Academic Publishing Workshop

We invite you to join MDPI for an Academic Publishing Workshop, where we will provide a detailed overview of MDPI's open access publishing model, covering important topics such as selecting the right journal, understanding the editorial process, and adhering to key principles of publication ethics. Attendees will also receive practical tips on enhancing their scientific writing and responding to reviewer feedback. Led by experienced Assistant Editor, Antigona Ulndreaj, this session is designed to equip participants with the knowledge and skills needed to strengthen their publication efforts.

Building A Level 3 A316

Enhancing Qualitative and Quantitative Text Analysis with Transparent Generative AI

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Enhancing Qualitative and Quantitative Text Analysis with Transparent Generative AI

Normand Peladeau, Provalis Research, Montreal, QC, Canada

This workshop introduces the integration of GenAI features in WordStat text analytics and QDA Miner qualitative analysis software, showing how generative methods can be combined with traditional approaches for more robust insights. We will highlight how high transparency was achieved to enhance reliability and trust. Participants will explore options for choosing AI engines, working with visible prompts, and integrating user-created or predefined scripts. Examples will range from simple tasks such as sentiment analysis, translation, text extraction, and preprocessing, to more advanced applications, including the fully automated analysis of open-ended responses and customer feedback—while still offering the flexibility to keep the human in the loop when interpretation and judgment matter most.

Saturday, October 25, 5:30 PM - 6:45 PM

Building A Level 3 A303

From Classroom to Industry: Modern Optimization with AMPL for Energy, Finance, Supply Chain, and Beyond

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - From Classroom to Industry: Modern Optimization with AMPL for Energy, Finance, Supply Chain, and Beyond

Marcos Dominguez Velad, AMPL, Plasencia, Spain, Gleb Belov

Optimization has never been more critical – or more accessible. This talk will show how students and applied researchers can bridge the gap between academic learning and real-world impact using AMPL in modern workflows. With free academic licenses now including commercial-grade access to leading linear solvers (Gurobi, CPLEX, Xpress, Mosek, COPT, HiGHS and more), AMPL offers an unmatched opportunity to develop and apply optimization skills directly in industry contexts. We will explore how AMPL integrates seamlessly with the tools you already know – Python, R, Jupyter, Colab, VS Code, cloud platforms, and modern data pipelines – making it a “set it and forget it” system for building models that stay faithful to the real problems you’re solving. Attendees will see how AMPL powers large-scale, mission-critical applications in energy, finance, transportation, and supply chain, enabling companies to save costs, increase efficiency, and improve decision-making. By combining intuitive modeling with cutting-edge solver technology, AMPL empowers students and researchers not just to learn optimization, but to contribute immediately to high-impact, industry-grade projects. Join us to discover how AMPL can position you, or your students, at the forefront of applied optimization.

Building A Level 3 A315

Advancing Optimization Excellence: How FICO® Xpress's Latest Features Scale Decision Intelligence

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Advancing Optimization Excellence: How FICO® Xpress's Latest Features Scale Decision Intelligence

Michael Perregaard, FICO, Birmingham, United Kingdom, Tristan Gally, Susanne Heipcke, Dinakar Gade, Carlos Zetina, Alexander Biele

Join our workshop to discover the cutting-edge innovations in Xpress Solver 9.8 and our updated APIs! Get a firsthand look at how Xpress Insight empowers business users to harness the full potential of analytical models faster than ever before. Experience the latest features and performance enhancements in FICO® Xpress Solver, including groundbreaking advances to our mixed-integer linear solver, the latest enhancements to our primal-dual hybrid gradient solver, and the newest features of FICO Xpress Global solver. We will present advancements in the Solver APIs and Xpress Mosel, outlining recent updates to modelling capabilities. In the second half of this presentation, we showcase how to accelerate Optimization Workflows with Xpress Workbench & Insight using Docker. We will demonstrate the ease of setup and minimal configuration required to achieve a unified development experience across Workbench and Insight using automated Docker-based workflows. Hear from Rusty Burlingame, Principle Operations Research Advisor of Southwest Airlines, speak about how Southwest is leveraging advanced analytics to improve customer experience, enhance employee engagement, and streamline operations to maximize revenue and support growth.

Building A Level 3 A316

Hexaly, a New Kind of Global Optimization Solver

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Hexaly, a New Kind of Global Optimization Solver

Fred Gardi, Hexaly, Brooklyn, NY, United States

Hexaly is a new type of global optimization solver. Its modeling interface is nonlinear and set-oriented. It also supports user-coded functions, enabling seamless integration of simulation with optimization or machine learning with optimization. The Hexaly API unifies modeling concepts from mixed-integer programming, nonlinear programming, and constraint programming. Under the hood, Hexaly combines various exact and heuristic optimization methods: spatial branch-and-bound, simplex methods, interior-point methods, propagation, automatic branch-cut-price, local search, and surrogate modeling. Hexaly stands out from traditional solvers like Gurobi, CPLEX, and OR-Tools by delivering super-fast solutions to problems such as routing, sequencing, scheduling, packing, clustering, matching, assignment, and location. For example, Hexaly provides solutions that are close to the best-known results in the literature for vehicle routing problems with thousands of points and scheduling problems with millions of tasks, achieving this in just one minute of runtime on a basic computer. In addition to the Optimizer, Hexaly offers an innovative platform called Hexaly Studio, designed to prototype, develop, and deploy optimization applications quickly in a low-code fashion. This web-based platform is also well-suited for educational purposes; like all other Hexaly products, it is free for faculty and students.

Building A Level 3 A304

Littlefield 2.0: A Factory Management Simulation for Teaching Operations Management

Vendor Workshop

Vendor Workshop

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Littlefield 2.0: A Factory Management Simulation for Teaching Operations Management

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

Littlefield for Operations is a competitive online simulation of either a factory or a medical laboratory that has been used by more than half a million students in 500+ universities in 60+ countries. Professors use the simulation as a pair of graded assignment to engage and excite and students in studying operations management topics like process analysis and inventory control. This presentation will introduce Littlefield 2.0 and allow participants to play a very fast and challenging scenario in the new student interface. Participants are encouraged but not required to bring a laptop or tablet to play the game.

Sunday, October 26, 8:00 AM - 9:15 AM

SA01

Building A Level 3 A301

Daniel H. Wagner Prize

Award Session

Daniel H Wagner

Chair: Konstantina Mellou, Microsoft Research, Redmond, WA, United States

1 - Analytics for Urban Bike Infrastructure Planning: A Multi-Year Collaboration with the City of Toronto

Bo Lin, University of Toronto, Toronto, ON, Canada, Madeleine Bonsma-Fisher, Shoshanna Saxe, Timothy Chan

Cycling promotes healthy living, reduces traffic, and combats climate change, but safety and comfort concerns remain key barriers to cycling uptake globally. In collaboration with the City of Toronto, we developed a suite of descriptive, predictive, and prescriptive analytics tools to support data-driven planning of urban bike infrastructure. As cities strive to promote sustainable transportation solutions, our work offers a user-friendly, scalable, and adaptable model for leveraging analytics to achieve these goals, paving the way for future research and collaboration in this critical area.

2 - Deepstock: Inventory Management with Policy Regularizations

Will Ma, Columbia Business School, New York, NY, United States, Xinru Hao, Jiayi Liu, Yaqi Xie, Linwei Xin, Lei Cao, Yidong Zhang

Deep Reinforcement Learning (DRL) is hard to deploy off-the-shelf for OR problems, due to imprecise data and computational difficulties. To overcome these issues, we show how to regularize DRL using well-established policy structures from OR, in particular for inventory management. We extensively analyze regularized vs. off-the-shelf DRL, demonstrating the power of policy regularizations when we understand the structure of near-optimal policies. We deploy our regularized DRL algorithm to replenish inventory at Alibaba's Tmall Supermarket, observing Pareto improvement across inventory metrics.

SA02

Building A Level 3 A302

Promoting Analytics, OR/MS, and Industrial Engineering to Undergraduate Students - Experience and Aspirations

Panel Session

Committee's Choice

Co-Chair: Ken Murphy, UC Irvine, Irvine, CA, United States

1 - Moderator Panelist

Ken Murphy, UC Irvine, Irvine, CA, United States

2 - Panelist

James Cochran, The University of Alabama, Tuscaloosa, AL, United States

3 - Panelist

Ana Muriel, University of Massachusetts Amherst, Amherst, MA, United States

4 - Panelist

Mihai Banciu, Bucknell University, Oxford, United Kingdom

5 - Panelist

Michael Veatch, Gordon College, Wenham, MA, United States

6 - Panelist

Cindy Evans, Amazon, Atlanta, GA, United States

SA03

Building A Level 3 A303

Innovations in AI And Machine Learning for Impact

Panel Session

Practice Track

Co-Chair: Raghu Kulkarni, Equifax, Alpharetta, United States

1 - Moderator Panelist

Raghu Kulkarni, Equifax, Alpharetta, GA, United States

As artificial intelligence and machine learning continue to evolve, so does their potential to drive transformative change across industries. This panel explores cutting-edge innovations that move beyond experimentation to deliver real-world impact. Featuring expert insights from Raghu Kulkarni and fellow leaders in data science, we will examine the intersection of emerging techniques, scalable deployment, and ethical responsibility. Key topics include next-generation model development, responsible AI, cognitive computing, and frameworks for ensuring measurable outcomes. Join us for a thought-provoking conversation on how organizations can harness innovation in AI/ML not just for advancement, but for lasting value.

2 - Panelist

Preetika Madan, Citi Group, Atlanta, GA, United States

3 - Panelist

Warren Hearn, OptiML AI, Atlanta, GA, United States

SA04

Building A Level 3 A304

Frontiers in Data-Driven Finance

Invited Session

Finance

Chair: Kay Giesecke, Stanford University, Stanford, CA, United States

Co-Chair: Junting Duan, Stanford University, Stanford, CA, United States

1 - Bi-Causal Optimal Transport and Application in Financial TIME Series Generation

Renyuan Xu, New York University, New York, NY, United States

In this talk, we present a foundational framework for bi-causal optimal transport between Markov processes defined on general state spaces. Building on this foundation, we develop a gradient-based learning algorithm grounded in the Dynamic Programming Principle, along with a novel policy gradient theorem tailored to the bi-causal setting. This framework is particularly suited for generating financial time series, especially in challenging scenarios involving missing data or heterogeneous data observation frequencies.

2 - Estimation of Out-of-Sample Sharpe Ratio for High Dimensional Portfolio Optimization

Weichen Wang, The University of Hong Kong, Hong Kong, Hong Kong, Xuran Meng, Yuan Cao

Portfolio optimization aims at constructing a realistic portfolio with significant out-of-sample performance, which is typically measured by the out-of-sample Sharpe ratio. However, due to in-sample optimism, it is inappropriate to use the in-sample estimated covariance to evaluate the out-of-sample Sharpe, especially in the high dimensional settings. In this paper, we propose a novel method to estimate the out-of-sample Sharpe ratio using only in-sample data, based on random matrix theory. Furthermore, portfolio managers can use the estimated out-of-sample Sharpe as a criterion to decide the best tuning for constructing their portfolios. Specifically, we consider the classical framework of Markowitz mean-variance portfolio optimization under high dimensional regime of $p/n \rightarrow c \in (0, \infty)$, where p is the portfolio dimension and n is the number of samples or time points. We propose to correct the sample covariance by a regularization matrix and provide a consistent estimator of its Sharpe ratio. The new estimator works well under either of the following conditions: (1) bounded covariance spectrum, (2) arbitrary number of diverging spikes when $c < 1$, and (3) fixed number of diverging spikes with weak requirement on their diverging speed when $c \geq 1$. We can also extend the results to construct global minimum variance portfolio and correct out-of-sample efficient frontier. We demonstrate the effectiveness of our approach through comprehensive simulations and real data experiments. Our results highlight the potential of this methodology as a useful tool for portfolio optimization in high dimensional settings.

3 - Imputation-Powered Inference for Missing Covariates

Junting Duan, Stanford University, Stanford, CA, United States

A prevalent problem in empirical research is missing covariate data as an input for model estimation. This paper develops a novel framework for optimally using partially observed covariates data for estimation and inference in a downstream model. Commonly used approaches, that either discard all partially observed samples or naively treat imputed values as real observations, are generally inefficient or biased. Our novel approach combines a bias correction with an optimal weighting scheme for imputed values that balances the efficiency trade-off between imputation error and effective sample size. Our method ensures valid inference while improving statistical efficiency by leveraging all available data. Our framework accommodates general missing data patterns for large panels and can be combined with a broad class of imputation methods. We establish the asymptotic normality of the proposed estimator under general assumptions. In simulations and an empirical application, we demonstrate the superior performance of our method over the naive approaches, as it achieves both lower bias and variance, while being robust to the imputation quality.

4 - Risk Disclosure Premium in the Cross-Section of Stock Returns: A Generative Topic Modeling Approach

Steve Yang, Stevens Institute of Technology, Hoboken, NJ, United States, Beichen Zhang, Shanshan Yang

We examine the pricing implications of financial risk disclosures by applying generative topic modeling to Section 1A of 10-K risk factor filings for S&P 1500 firms from 2010 to 2023. Using embeddings from FinBERT, GTE-Qwen, and SFR-Mistral, we quantify firm-level exposure to specific risk narratives and compute corresponding risk scores. These scores are used to construct narrative-based risk factors through a two-step empirical asset pricing framework. First, we test the return predictiveness of these factors using the Fama-French 5 factor model via long-short portfolio sorts. Second, we implement Fama-MacBeth cross-sectional regressions to evaluate their risk premia alongside the standard market, size, value, profitability, and investment factors. Our results reveal a robust risk premium in operational risk disclosure - which remains statistically significant even after applying the Bonferroni correction for multiple tests. In addition, several other disclosure factors, such as funding risk, intellectual property risk, and infrastructure risk, also show some level of significance. These findings suggest that firm-disclosed forward-looking risks, as identified through generative embedding model on firm risk disclosures, offer economically meaningful and statistically robust insights into cross-sectional return variation.

SA05

Building A Level 3 A305

HAS Junior Faculty Lightning Session

Panel Session

Junior Faculty Interest Group

Co-Chair: Zhaowei She, Singapore Management University, Singapore, Singapore

Co-Chair: Xiaoquan Gao, Singapore Management University, Singapore, Singapore

1 - Moderator Panelist

Xiaoquan Gao, Singapore Management University, Singapore, Singapore

2 - Moderator Panelist

Zhaowei She, Lee Kong Chian School of Business, Singapore Management University, Singapore, Singapore

3 - Panelist

Yanhan (Savannah) Tang, SMU Cox School of Business, Dallas, TX, United States

4 - Panelist

Jiatao Ding, Nanyang Technological University, Singapore, Singapore

5 - Panelist

Yiwen Jin, Haskayne School of Business, University of Calgary, Calgary, AB, Canada

6 - Panelist

Jingyuan Hu, UCLA Anderson School of Management, Los Angeles, CA, United States

7 - Panelist

Yu Ma, Massachusetts Institute of Technology, Cambridge, MA, United States

8 - Panelist

Raghav Singal, Tuck School of Business at Dartmouth, Hanover, NH, United States

9 - Panelist

Yiwen Dong, University of Illinois, Urbana-Champaign, Santa Clara, CA, United States

10 - Panelist

Zhaohu(Jonathan) Fan, Scheller College of Business, Georgia Institute of Technology, Atlanta, GA, United States

11 - Panelist

Wilson Lin, Santa Clara University, Santa Clara, CA, United States

12 - Panelist

Emily Fainman, Texas State University, San Marcos, TX, United States, Beste Kucukyazici, TING WU

13 - Panelist

Jingyuan Hu, UCLA Anderson School of Management, Los Angeles, CA, United States

14 - Panelist

Yvonne Huijun Zhu, National University of Singapore, Singapore, Singapore

15 - Panelist

Yuwei Zhou, Indiana University, Bloomington, IN, United States

16 - Panelist

Xiaoquan Gao, Singapore Management University, Singapore, Singapore

17 - Panelist

Adam Meyers, University of Miami, Miami, FL, United States

18 - Panelist

Isabelle Rao, University of Toronto, Toronto, ON, Canada

SA06

Building A Level 3 A306

INFORMS Case Competition

Award Session

Case Competition

Chair: A. Cecilia Zenteno, Massachusetts General Hospital, Boston, MA, United States

1 - Calyber: a Ridesharing Game

Chiwei Yan, University of California Berkeley, San Francisco, CA, United States, Yifan Shen, Julia Yan

Motivated by the persistent challenges that ridesharing platforms face in offering shared rides (a service that pools riders into a single vehicle), this case offers a hands-on and engaging experience in developing real-time analytics solutions. Using a rich historical ridesharing dataset, students design and test policies for quoting prices and matching riders, competing for top performance on a test set. By completing this case, students gain a practical understanding of stochastic dynamic optimization, a concept often difficult to teach through traditional instruction.

2 - Reinforcement Learning for Clinical Decision Support

Dessislava Pachamanova, Babson College, Wellesley, MA, United States, Angela Lin, Georgia Perakis

This case, based on a project with a U.S. academic medical center, addresses real-world challenges in developing AI-augmented decision support for patient treatment. It highlights the application of reinforcement learning to personalize care, contrasting it with a static segmentation approach. The case covers problem framing, data preprocessing, supervised/unsupervised/reinforcement learning, stakeholder engagement, and AI ethics. Python code and realistic data are included.

3 - Routing for Social Good: Optimizing Human Milk Bank Logistics

Natalia Summerville, Memorial Sloan Kettering Cancer Center, New York, NY, United States, José Antonio Carbajal

This case challenges students to optimize donor milk collection for a human milk bank using real-world data. Students develop routing and scheduling models to minimize travel time while ensuring timely pickups. Variations of the case introduce predictive time series models to forecast donation volumes, enabling smarter collection plans. Ideal for analytics, OR, or supply chain courses, the case blends logistics, forecasting, and public health impact.

SA07

Building A Level 3 A307

Strategic Considerations in Transportation

Job Market Showcase

Job Market Showcase

Chair: Kyle Bathgate, The University of Texas at Austin, Austin, 78751, United States

1 - Bridging Transportation Infrastructure Design, Operational Policies, and User Interactions: a Mathematical Framework for Curb Space Management

Jisoon Lim, University of Michigan, Ann Arbor, MI, United States

Today's transportation systems are undergoing transformative change, driven by advances in connectivity, autonomy, shared mobility, and electrification. These innovations create new opportunities for enhanced mobility and multimodal integration but also call for a fundamental rethinking of infrastructure design and operational strategies to ensure accessibility, efficiency, and equity. My research addresses this challenge by focusing on the dynamic interplay among three interdependent pillars: infrastructure design, operational policies, and user interactions. I develop mathematical frameworks that integrate these pillars, capturing their interactions while balancing model efficiency and granularity. This tripod approach offers structured and scalable tools for analyzing transportation operations that are both socially responsive and practically implementable. I highlight curb space as a key example of this evolving landscape, where demand has grown in both diversity and intensity. As a vital interface for urban activities, curb space must accommodate both traditional and emerging mobility needs, providing equitable and proximate access to key urban functions. My research employs mathematical solutions - drawing from optimization, scalable algorithms, game theory, simulation, and behavioral science - to construct technology-enabled infrastructure models, design efficient operational policies, analyze user behaviors through data, and develop solution algorithms and software tools that capture the dynamic nature of curb space. These approaches aim to identify socially optimal solutions and guide user adaptation to emerging curb management strategies. By bridging technology, operations, and behavior, my work contributes to the design of integrated, user-centered, and future-ready mobility systems that can serve as models for other societal infrastructure challenges.

2 - Equitable Transit Design in the Era of Telework

Tianxing Dai, Northwestern University, Evanston, IL, United States, Marco Nie

Transit ridership in the US has plunged due to the COVID-19 pandemic and has yet to recover to the pre-pandemic level primarily due to the wide adoption of telework. Not only do previous transit commuters choose to telework a few days a week, but they also tend to drive more than before on the days they do commute. For transit agencies, the structural reduction of transit ridership means a loss in the farebox revenue needed to sustain their services. When the COVID-19 relief funds expire, transit agencies need to seek other means to cover the budget deficit. If they end up with cutting service, the disadvantaged population who rely heavily on public transit would be disproportionately affected. To help transit cope with telework, we developed a microeconomic model of transit design under the monocentric city setting, and endogenized travelers' commute mode (driving, transit or telework) and housing location choices, as well as traffic congestion. To understand the equity implications of telework on transit, our model features heterogeneous travelers who have different income levels. The transit design model permits adjustments in transit services and finance policies including fare, tax and congestion charges, and we adopt both a utilitarian objective and a Rawlsian objective to study the efficiency-equity trade-off. When analyzing and solving the location-mode choice equilibrium, we capture both the spatial mode sorting effect and the income mode sorting effect, where residents living farther from the city center and with higher income are more likely to telework.

3 - The Transportation Network Efficiency Effects of Congestion Pricing

Huahua Zhu, Indiana University, Bloomington, IN, United States

Traffic congestion remains a persistent challenge in modern urban areas, contributing to longer travel times, increased travel costs, higher gasoline consumption, and network inefficiency. To evaluate the impacts of congestion pricing on mitigating these negative externalities, this paper develops a structural decision-making framework that models agents' optimal travel route and mode choice behavior under uncertainty, accurately capturing the resulting traffic equilibrium conditions. Within this framework, two user optimization problems are formulated—one without congestion pricing and one incorporating it—allowing for the establishment of equilibrium existence and uniqueness, as well as the derivation of closed-form expressions of the equilibria for both cases. Additionally, a system optimization problem is introduced from a social planner's perspective, providing a benchmark for evaluating network efficiency in response to transportation policy interventions. Empirical evidence from New York City's cordon-based congestion pricing policy demonstrates that such interventions significantly alter traveler distribution patterns, reduce travel times and travel costs, and lower gasoline consumption at the main entry links. However, this policy only improves the overall network efficiency by 4% within a monocentric city structure. Furthermore, an ex-ante policy analysis suggests that increasing the congestion pricing from 9 dollars to 12 dollars would reduce the network efficiency improvement to 3%.

4 - Methods for Assessing the Resilience of Maritime Freight Transportation Systems

Kyle Bathgate, The University of Texas at Austin, Austin, TX, United States

Resilient systems are critical for business continuity and effective supply chain management. In this work, new methods for measuring the resilience of maritime freight infrastructure facilities are developed. The metrics are derived using anchorage queue lengths and wait times, and relate to the system robustness and recoverability. Applications of the metrics are presented using historic vessel tracking automatic identification system (AIS) data at the Port of Houston, Texas to demonstrate the resilience of the system to disasters such as hurricanes and recurring fog events.

When coupled with discrete-event simulation approaches, the metrics allow for the testing of hypothetical disasters and scenario planning, allowing stakeholders to observe the potential resilience implications of a given infrastructure investment or operational adjustment. These findings are relevant to public and private sector maritime industries as they allow for increased awareness of system response under distress and provide tools for increased supply chain resilience.

SA08

Building A Level 3 A310

Managing Emerging Technologies: AI, Blockchain, and Digital Markets

Invited Session

Information Systems

Chair: Xue Tan, Southern Methodist University, Dallas, United States

1 - Resalable Bytes: Unlocking Digital Ownership in Content Markets Through Blockchain

Zixuan Meng, The University of Texas at Dallas, Richardson, TX, United States, Yi Gao, Ming Fan

The proliferation of non-fungible tokens (NFTs) has sparked both public skepticism and institutional interest. While critics dismiss NFTs as speculative assets, leading firms increasingly adopt NFT-based strategies that promise transferable digital ownership, particularly in markets historically dominated by licensing restrictions and Digital Rights Management (DRM). Our work develops an analytical model to investigate how blockchain-enabled ownership, through granting resalable rights, reshapes digital market dynamics. Specifically, we compare three distinct market architectures: (1) a benchmark model without ownership, (2) a creator-initiated blockchain (CIB) model, and (3) a platform-initiated blockchain (PIB) model. Our findings show that both blockchain models enhance consumer surplus and social welfare when initiator's operating costs and user adoption frictions are relatively low. However, who initiates still matters, leading to distinct (1) changes in equilibrium price, (2) shifts in market demand, and (3) the specific channels through which ownership value is realized. This asymmetry between creator- and platform-initiated models is also reflected in their incentives. Creators are more likely to initiate blockchain, even under high operating and consumer adoption costs. This asymmetry in incentives leads to strategic responses by platforms, which may preemptively adopt blockchain despite short-term revenue losses, in order to mitigate long-term disadvantages associated with being a second mover.

2 - Civic Monitoring Through Internet-of-Things

Brian Lee, Pennsylvania State University, University Park, PA, United States

Internet-of-Things (IoT) devices are technological artifacts that integrate sensors, data, and automation to support decision-making. While widely adopted in areas such as marketing, manufacturing, and logistics, IoT is increasingly being used to promote community safety. However, its effectiveness remains uncertain due to the longstanding presence of traditional regulatory monitoring practices. This study focuses on consumer-grade air quality sensors as a form of IoT and examines whether their adoption can enhance community safety by reducing air pollution, even in areas already served by EPA air monitoring stations.

3 - AI Algorithms and Collusion in Ad Auctions

Subrahmanyam Aditya Karanam, National University of Singapore, Singapore, Singapore, Ashish Agarwal, Anitesh Barua

The rapid growth of Artificial Intelligence (AI) has led to its widespread deployment in automating decision-making and growing concerns about tacit collusive behavior among AI algorithms. One application involves real-time bidding for display advertisements, where collusion among bidders can negatively impact auctioneer revenues. Our study investigates this issue by analyzing the behavior of budget-constrained deep Q-learning algorithms in First-price auctions (FPA) through simulated computational experiments. We also examine how the advancement of AI technology and data policies affect the competitiveness of the auction. Overall, we find collusion among algorithmic bidders to be prevalent in FPA. The competitiveness of auctions further diminishes with more advanced AI algorithms. Contrary to findings in the literature on AI bidding with no budget constraints, our results suggest that when advertising platforms provide more detailed information about the auctions, it facilitates collusive behavior among budget-constrained AI bidders. Our results highlight the impact of using AI-powered algorithms on the outcomes in FPA and call for revisiting their design.

4 - Sell Data to AI Algorithms Without Revealing It: Secure Data Valuation and Sharing via Homomorphic Encryption

Michael Yang, University of Texas at Dallas, Richardson, TX, United States, Ruijiang Gao, Zhiqiang Zheng

Traditional data-sharing practices requires data owners to reveal the data to buyers to determine its value before they can negotiate a fair price, creating legal exposure, privacy risk, and asymmetric information that discourages exchange. We propose a homomorphic-encryption (HE) framework that enables prospective buyers to quantitatively assess a dataset's utility to AI algorithm while the data remain fully encrypted end-to-end. This approach tackles the last-mile problem in building modern AI data marketplaces: Efficient Encrypted Utility Estimation. We designed a lightweight data utility evaluation method using HE protocols that allows buyers to score different data samples without actually having to obtain the raw data. The proposed method can work with popular gradient-based data valuation methods and can scale to large foundational models such as Large Language Models (LLMs). By allowing organizations to determine the value of their data, without disclosing the data itself before the transaction, our work provides a practical path toward safer data monetization in modern AI data marketplaces.

5 - The Impact of Generative AI on Open-Source Community Engagement

Karthik Babu Nattamai Kannan, Southern Methodist University, Allen, TX, United States, Narayan Ramasubbu

Emerging research highlights significant productivity gains from generative AI tools like GitHub Copilot in software development. However, the specific mechanisms behind these gains, especially in distributed development settings, remain underexplored. This study examines GitHub's trial of "Copilot for Pull Requests (PR)" to assess its effect on PR latency—the time taken to close PRs. Analyzing over 100 open-source repositories and approximately 200,000 PRs using Propensity Score Matching and panel data analysis, we find that Copilot usage reduces PR latency by up to 4 days on average. This improvement is primarily due to a reduced need for multiple reviewers, streamlining the review process. Repository-level data shows that higher Copilot adoption leads to cumulative latency reductions, highlighting its scalable impact. Further investigation suggests that Copilot-generated summaries enhance collaboration by providing precise and comprehensive documentation at the time of PR creation, decreasing the number of review iterations. These findings offer key insights into how AI can optimize collaborative software workflows. Copilot not only boosts individual productivity but also mitigates systemic inefficiencies in the pull-based development model. The study underscores the value of integrating AI tools into the software development lifecycle and highlights the need to address potential challenges such as workflow friction and information gaps. This research advances understanding in AI-driven productivity and provides actionable guidance for practitioners aiming to improve project performance through AI-enabled development tools.

SA09

Building A Level 3 A311

Novel Empirical and Experimental Analysis

Job Market Showcase

Job Market Showcase

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

1 - AI-Driven Predictive Analytics for Strategic Decision-Making in Industry and Public Health

Mahima Naznin, University Of Missouri- Columbia, Columbia, MO, United States

In the industrial domain, a predictive framework was developed in one project for forecasting steel prices, which are notoriously volatile due to fluctuating global market conditions. Through rigorous correlation and covariance analysis, novel economic indicators were identified that influence price shifts. These insights were integrated into a machine learning-enhanced time series model—capable of predicting steel prices up to 12 weeks ahead with high accuracy. The model was operationalized in an interactive dashboard to support procurement strategies, enabling data-driven decisions that mitigate financial risk and improve supply chain efficiency.

In another project in the healthcare domain, the critical need was addressed for early detection of Type 2 Diabetes (T2D) risk among adolescents. A machine learning-based screening tool was developed, incorporating socioecological, behavioral, psychological, and physiological factors. After benchmarking multiple algorithms, XGBoost demonstrated the strongest performance with an AUC-ROC of 0.72, enabling early identification of high-risk individuals up to 15 years in advance.

Together, these projects illustrate my ability to build end-to-end, impactful AI-driven solutions—from feature discovery to model deployment—in high-stakes environments. My work emphasizes practical implementation, interpretability, and domain relevance, contributing to smarter decision-making in both business and health sectors.

2 - Decoding Cognitive Aging

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

Accurately identifying individuals at risk of cognitive decline and those who demonstrate exceptional cognitive resilience is critical for advancing early intervention strategies and understanding the neural basis of healthy aging. This work develops and evaluates computational frameworks to distinguish between cognitive aging trajectories using neuroimaging and longitudinal cognitive data. Specifically, it focuses on characterizing individuals with superior cognitive performance, referred to as "Positive-Agers", in contrast to their age-matched peers exhibiting cognitive decline (i.e., "Cognitive Decliners"). Drawing on large-scale datasets from the UK Biobank, two algorithms are introduced: Optimal Labeling with Bayesian Optimization (OLBO), which integrates neuronal functional connectivity networks; and Optimal Cognitive Scoring (OptiCS), a predictive pipeline that leverages diffusion MRI to assess and distinguish cognitive groups. Both methods demonstrate strong performance (AUC 83–88%) and identify key neuroanatomical substrates associated with cognitive resilience. In addition, a labeling mechanism is proposed for cognitive assessment and monitoring. These findings enhance early detection of cognitive decline and offer insights into the neural correlates of cognitive aging.

3 - Drivers of Customers' Preferences for AI Service Agents

Chia-Chun Yang, University of Cincinnati, Cincinnati, OH, United States, Craig Froehle, Yinghao Zhang

The use of artificial intelligence (AI) in customer service roles is evolving rapidly across diverse industries such as finance, retail, education, and healthcare. Leveraging advanced language processing algorithms, AI-powered service agents (AISA) are increasingly capable of engaging in service tasks such as checking order statuses, facilitating food orders, recommending personalized items, and providing mental health support. Additionally, recent studies suggest that AISA has the potential to outperform human service agents not just in efficiency and consistency, but also potentially empathy and other overtly "human" traits. However, despite their growing capabilities and many associated benefits, AISA still faces challenges in customer acceptance. Much is still unknown about the determinants of customers' preference for using AISA, and how much this technology shapes the customer service experience. A better understanding of these issues can enable organizations to make substantive operational advancements: improve service process designs, increase efficiency, and enhance customer satisfaction. This study investigates two research questions: (1) How do trust and perceived performance risk influence customers' preference for AI-powered service agents? (2) What factors drive customers' perceptions of human-like trust and perceived performance risk? Specifically, we examine how these perceptions are influenced by two factors: the complexity of the customer's task and the degree of anthropomorphism embedded in AI-powered service agents. Experimental methods are conducted over multiple studies to gain insights into these questions.

SA10

Building A Level 3 A312

Markets, Risk, and Regulation

Job Market Showcase

Job Market Showcase

Chair: Ramin Sepehrirad, Washington State University, PULLMAN, WA, United States

1 - Robust Optimal Strategies for Two-Period Liquidation in Financial Systems

Hongyi Jiang, The Chinese University of Hong Kong, Hong Kong, China, People's Republic of

We study the problem of asset liquidation in financial systems. During financial crises, asset liquidation is often inevitable but can lead to substantial losses if a significant amount of illiquid assets are sold simultaneously at depressed prices. To tackle this challenge, we consider a two-period liquidation model that allows for preemptive liquidation before maturity and propose a worst-case approach to the associated liquidation problem. Specifically, we propose a robust optimal strategy—a tractable liquidation approach that maximizes the worst-case terminal value of liquid assets, taking into account the uncertainty of other banks' liquidation decisions. In addition, we find that the unique Nash equilibrium is attained when all banks adopt our proposed strategy. We further demonstrate that the robust optimal strategy retains a similar form even when interbank liabilities are involved, where we consider both scenarios of full and partial network information. While our analysis builds upon a stylized model, our findings offer valuable guidelines for developing robust liquidation strategies that mitigate losses resulting from asset liquidation.

2 - On Gpo Changes Within the Healthcare System

Andrew Hildebrandt, Auburn University, Auburn, AL, United States

Within our healthcare network, a significant transformation is underway—hospitals are rapidly shifting their procurement strategies in a dynamic and increasingly cyclical pattern. Rather than maintaining long-standing affiliations, many hospitals are now engaging in a process of joining, dropping, and even rejoining group purchasing organizations (GPOs). This behavior reflects a broader reevaluation of how hospitals manage supply chain efficiency, cost control, and strategic sourcing in response to evolving external pressures.

The traditional model of stable, long-term GPO participation is giving way to a more flexible and performance-sensitive approach. Hospitals are no longer viewing GPO membership as a static commitment but as a strategic lever that can be adjusted based on changing market conditions, policy environments, and internal performance metrics. Some hospitals drop out in pursuit of direct contracting opportunities or greater autonomy, only to rejoin a GPO later when the benefits—such as aggregated buying power or access to specific supplier networks—become more compelling again.

This cyclical pattern of GPO engagement signals a move toward more adaptive procurement behavior. It also reveals how hospitals are leveraging procurement decisions not only to reduce costs but also to respond to new operational challenges, shifts in payer mix, regulatory uncertainty, and technological change. Ultimately, this loop of entry, exit, and re-entry into GPO arrangements underscores the increasing complexity and strategic importance of supply chain management in healthcare today.

3 - Taking My Time to Perfect the Beat: A Behavioral Analysis of Product Launch Delay Decisions

Hyunsuk Baek, Arizona State University, Tempe, AZ, United States, Huseyn Abdulla, Seongkyoon Jeong

A product launch is a crucial event in the new product development process, and the timing of a product launch is a crucial decision in many innovation-driven industries. More often than not, product launches are delayed. The impetus behind many well-publicized product launch delays has been the aspiration of achieving high product quality given ever-increasing market expectations. Even products with pre-announced launch dates are often delayed to accommodate continued or final quality improvements. However, by delaying product launches beyond pre-announced dates, managers and entrepreneurs face risks associated with shifting market expectations due to technological innovations and new competitive entries, as well as additional development costs with typically diminishing returns. Recognizing these trade-offs, we conduct a behavioral analysis of decision-making regarding a product launch delay beyond a pre-announced launch date. We develop an analytical model and derive a rational benchmark for launch delay decisions. Then, we conduct a behavioral experiment to test the extent to which individuals' actual delay decisions are aligned with this rational benchmark. We find that most individuals largely deviate from the rational benchmark, exhibiting what we term as under-delay and over-delay biases. This results in, on average, 24.25% of relative profit loss to the rational benchmark. Given this salience, we propose a behavioral model to explain the launch delay decisions. Estimating the behavioral model, we find that a delay-dependent status quo effect and loss-aversion effect together can explain individuals' decisions in the experiment.

4 - Regulatory Intervention in Automotive Recalls

Ramin Sepehrirad, Washington State University, PULLMAN, WA, United States, Kevin Mayo, George Ball

Despite extensive research on automotive recalls, the actual effectiveness of remedial actions—how well firms repair faulty products—is largely unknown, even as millions of unrepaired defective vehicles pose significant safety risks on U.S. roads. We address this gap by examining how regulatory intervention by the National Highway Traffic and Safety Administration (NHTSA) impacts automakers' remedial performance. Using an attention-based view, we posit competing effects: NHTSA-initiated recalls may heighten managerial focus due to increased scrutiny, yet prolonged investigations and repair complexities associated with such recalls could divert decision makers' attention. Employing a quasi-experimental design on a dataset of over 4,200 U.S. passenger vehicle recalls (2009-2024), we surprisingly find that regulator-initiated recalls significantly reduce repair rates—by as much as 5%. We further explore explanatory mechanisms, such as recall complexity and repair procedures. Our findings offer crucial implications for NHTSA's policies and industry practices to improve recall outcomes and public safety.

SA11

Building A Level 3 A313

Scheduling and Staffing in Healthcare Applications

Job Market Showcase

Job Market Showcase

Chair: Yiqi Tian, University of Pittsburgh, Pittsburgh, PA, United States

1 - Stochastic Programming Approaches for Medical Residents' Shift Scheduling Under Residents Absenteeism and Surging Demand

Shutian Li, Lehigh University, Bethlehem, PA, United States, Karmel Shehadeh

Motivated by our collaboration with a general residency program, we present a new two-stage stochastic mixed-integer programming approach for resident shift scheduling, considering residents' absenteeism and demand uncertainty. We proposed an enhanced Logic-Benders-based decomposition algorithm to solve the proposed model. We also developed a user-friendly tool that implements the proposed approach. Numerical results demonstrate the computational efficiency of our proposed algorithm over state-of-the-art decomposition algorithms and provide valuable practical insight.

2 - Early Intervention for Chronic Pain: Reducing Healthcare Utilization, Surgeries and Healthcare Costs

Buyun Li, Kelley School of Business, Indiana University, Bloomington, IN, United States, Mohammad Zhalechian, Christopher Chen

Chronic pain imposes significant clinical and economic burdens, yet optimal care pathways remain contested. We estimate the causal impact of early interventional pain management (IPM) on healthcare utilization, surgeries, and costs using a natural experiment from a 2019 insurer policy eliminating co-pays for conservative care in select regions. Leveraging administrative claims data and a difference-in-differences instrumental variable approach, we find that early IPM reduces downstream costs, surgeries, and service use. These findings provide the first causal evidence supporting early IPM access as a high-value strategy and inform benefit design and payment models promoting efficient pain care.

3 - Primary Care Panel Optimization

Dilara Sonmez, The University of Chicago Booth School of Business, Chicago, IL, United States, Daniel Adelman, George Weyer

The method to achieve financial success under value-based contracts remains unclear, but recent studies suggest that tailoring patient visit patterns could be the key. With CMS aiming to have all Medicare beneficiaries in an accountable care relationship by 2030, optimizing primary care visit for patients is crucial for the success of this initiative for both provider organizations and health policymakers. Despite having evidence that primary care visit patterns and having a large primary care network are important to financial success under value-based contracts, no one has yet provided a framework to model the effect of visit patterns and to implement them as decision variables in an optimization problem in a realistic setting with resource constraints. We address this gap in the literature by formulating the problem of finding physicians' optimal patient panels and prescribing patients with a primary care dosage as a mixed-integer nonlinear optimization problem. Central to our method is a dynamic programming model that captures the physician's decision-making process for scheduling follow-up visits. Our approach also accounts for the complex dependencies

among key visit pattern metrics: frequency, regularity, and continuity of care. To ensure scalability, we introduce a technique that enhances computational efficiency in large-scale applications while maintaining solution quality.

4 - Modeling Human-Centric Healthcare Scheduling with Decision-Dependent Uncertainty

Yiqi Tian, University of Pittsburgh, Pittsburgh, PA, United States, Bo Zeng

This talk introduces decision-dependent uncertainty to capture human behavior in healthcare scheduling tasks, including blood drives and radiologist staffing. By human behavior in contexts, the proposed models enhance the reliability and robustness of operational plans, ultimately improving resource allocation and outcomes in these critical, human-centric decision processes within healthcare domain.

SA12

Building A Level 3 A314

Topics in Scheduling

Job Market Showcase

Job Market Showcase

Chair: Jorge Huertas, Georgia Tech - ISyE / AI4OPT, Atlanta, GA, United States

1 - Flow Shops with Reentry: a Theoretical and Computational Analysis

Nicklas Klein, University of Bern, Bern, Switzerland, Michael Pinedo

Scheduling manufacturing processes efficiently is a key challenge in many supply chains. We consider manufacturing processes in which the jobs traverse and reenter a flow shop multiple times, i.e., after a job has traversed the shop and completed its processing on the last machine, it must return to the first machine and traverse the shop again until it has completed all its required loops. Such a setting is referred to as a flow shop with reentry and has numerous industrial applications, e.g., in the electronics industry. The planning problem is to schedule all loops of all jobs so that an objective related to customer satisfaction or resource utilization is optimized. We propose several algorithms and analyze their theoretical properties and computational performance.

2 - Three Essays on Multiple-Orders-Per-Job Scheduling Problems in a Two-Stage Flow Shop

Rohan Korde, Arizona State University, Tempe, AZ, United States

Motivated by semiconductor manufacturing, in my PhD research, I study single-objective and bi-criteria job scheduling problems at a 300-mm wafer foundry fabrication (fab) facility. Customers send their orders to a fab in terms of number of 300-mm wafers. At a fab, the chip manufacturing involves hundreds of reentrant processing steps on many machines. The wafers are moved between machines using front opening unified pods (FOUPs or jobs) using an overhead hoist transportation (OHT) system. When the customer orders increase, the fab is constrained by the number of jobs available because by increasing the jobs can lead to congestion in the OHT. This can cause tardy orders, which the fab wants to minimize. One way to address this is to pack orders from different customers in the same job and then schedule the jobs to be processed on the machines. This class is called “multiple orders per job” (MOJ) scheduling problems. I model the fab as a two-stage permutation flow shop. In the first paper, I study single-objective MOJ problems (each to minimize a single regular performance measure) where the orders are released to flow shop at the same time. In the second paper, I study single-objective MOJ problems where the orders are released to the flow shop at different times (unequal release time). In the third paper I study bi-criteria MOJ problems. I use mixed integer linear programming, heuristics, and metaheuristics to solve these problems. I use full-factorial computational experiments to compare the solution quality.

3 - Safety-Driven Scheduling via Online Multi-Agent Reinforcement Learning

Pavithra Sripathanallur Murali, George Mason University, Fairfax, VA, United States

Scheduling in multi-component networks must balance performance with safety considerations, particularly in safety-critical systems. While existing decisions approaches either rely on model-based techniques like reachability analysis requiring precise system safety dynamics or implement model-free methods with simplistic binary safety measures and constraints falling short in capturing the gradients of safety. To this end, we propose a novel approach for decentralized scheduling in multi-component networks with interdependencies that incorporates continuous safety indicator measures for each schedule, enabling more nuanced safety assessment to classify and eliminate them during learning process. We specifically utilize decentralized markov decision process to model the decision process of network scheduling. Our method demonstrates superior performance in identifying and eliminating unsafe schedules while maintaining finite time guarantees. We validate our approach through numerical experiments and a real-world case study of a Water Distribution Network in Tampa, FL, where we optimize restoration scheduling under contamination risks. This work provides infrastructure managers with a practical framework for evaluating resilient strategies under simulated disruptions.

4 - Scheduling for Semiconductor Manufacturing

Jorge Huertas, Georgia Tech - ISyE / AI4OPT, Atlanta, GA, United States, Pascal Van Hentenryck

Recent disruptions in the semiconductor supply chain—such as the global chip shortage—highlight the urgent need for resilient scheduling tools within wafer fabrication facilities. These tools must incorporate a range of operational complexities, including the diverse batching regimes employed across different fab areas. This talk introduces novel constraint programming models for parallel and serial batching with incompatible job families, common in critical steps such as the diffusion and ion implantation stages in semiconductor manufacturing. Evaluated on standard testbeds, the proposed models outperform existing approaches from the literature. Furthermore, they effectively scale to replicated industrial instances, where exploiting the problem's structure further enables rapid schedule adjustments. Overall, these advancements offer a powerful foundation that paves the way for robust and adaptive planning under uncertainty.

SA13

Building A Level 3 A315

Machine Learning and Applied Probability Job Market Session

Job Market Showcase

Job Market Showcase

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

1 - When Worst-Case Isn't Robust: on the Limitations of Distributionally Robust Formulations in Secretary Problems

Anas Abdelhakmi, National University of Singapore, Singapore, Singapore, Andrew Lim

We formulate and solve a Distributionally Robust Optimization (DRO) version of secretary problems, revealing a surprising phenomenon: the optimal robust decision coincides exactly with the nominal decision across various secretary problem variants and for different uncertainty sets (e.g., Kullback-Leibler divergence, Wasserstein distance\dots). We show that this result emerges because genuine robustness against distributional misspecifications is fundamentally unattainable in secretary problems without altering the problem structure itself. We illustrate how in this case, traditional DRO approaches yield only an illusion of robustness, leaving decisions equally sensitive, or even more vulnerable to distributional shifts if decision-makers misinterpret the ‘robust’ label. To address these vulnerabilities, we propose ‘Hedging for Robustness’, a managerial framework enabling practitioners to achieve robustness beyond conventional worst-case strategies. More generally, these results offer new intuitions of what it means for sequential decision to be truly robust and show how practitioners can achieve robustness beyond mere worst-case strategies.

2 - On the Minimax Regret of Sequential Probability Assignment via Square-Root Entropy

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

We study the problem of sequential probability assignment under logarithmic loss, both with and without side information. Our objective is to analyze the minimax regret -- a notion extensively studied in the literature -- in terms of geometric quantities, such as covering numbers and scale-sensitive dimensions. We show that the minimax regret for the case of no side information (equivalently, the Shtarkov sum) can be upper bounded in terms of sequential square-root entropy, a notion closely related to Hellinger distance. For the problem of sequential probability assignment with side information, we develop both upper and lower bounds based on the aforementioned entropy. The lower bound matches the upper bound, up to log factors, for classes in the Donsker regime (according to our definition of entropy).

3 - Statistical Methods for Benchmarking Quantum Random Circuits**Tudor Manole, Massachusetts Institute of Technology, Cambridge, MA, United States**

Random circuit sampling (RCS) is a technique for benchmarking quantum computers that was popularized by Google Quantum AI, and formed the basis of their quantum advantage demonstration in 2019. This task consists of sampling bitstrings from a randomly-generated quantum circuit, and is considered to be intractable on a classical computer when the circuit size is sufficiently large. An important step for carrying out RCS experiments is to estimate the fidelity of the circuit, which is roughly the probability that it runs without errors. Typical approaches to fidelity estimation require knowledge of the probability distribution of bitstrings generated by the quantum circuit, which is usually obtained via classical computation for small system sizes. In this work, we instead consider obtaining side information in the form of bitstring samples from a clean quantum computer. We derive sharp bounds on the sample complexity of fidelity estimation, showing that it improves polynomially in the number of clean bitstring samples available. As a special case, we also obtain the sharp sample complexity of fidelity estimation in the absence of side information, which scales exponentially in the number of qubits. These sample complexity bounds are achieved by practical statistical estimators, which we illustrate on real RCS data.

This talk is based on joint work with Daniel Mark, Wenjie Gong, Bingtian Ye, Soonwon Choi, and Yuri Polyanskiy.

4 - Adaptive Learning to Prevent Spiral-Down in Branded-Fare Revenue Management**Lan Wu, National University of Singapore Business School, Singapore, Singapore, Changchun Liu, Chung Piau Teo, Xun Zhang**

Airline revenue management increasingly relies on branded fares, which organize ticket options into structured families with tiered services. Traditional RM models assume static, exogenous arrival rates and homogeneous choice behavior, neglecting dynamics such as competition-driven price shifts and context-dependent passenger decisions. A prominent example is the “buy-down” phenomenon, where customers qualified for premium fares opt for lower-tier options, triggering a spiral-down effect that erodes revenue.

Unlike “blind” RM—which treats every no-sale as an observed rejection and assumes a single choice model for all arrivals—our method infers both latent arrival intensity and heterogeneous choice parameters directly from the censored booking curve. We introduce a framework that combines a dynamically adjusted marginal-fare transformation with an online Bayes–Selector learning scheme. Under mild separation conditions on purchase probabilities and revenue differentials, the algorithm requires only an $O(\log T)$ exploratory phase to learn context-dependent parameters; if prior information or historical data already yield sufficiently accurate estimates, exploration is skipped and the regret is $O(1)$.

Crucially, by continuously updating estimates of each segment’s willingness-to-pay and buy-down propensity, this anticipatory pricing prevents customers from systematically trading down for minimal savings, thereby breaking the spiral-down feedback and preserving premium fare revenue. In the exploitation phase, LP-based bid-price controls optimize fare levels and seat allocations in real time. Extensive simulations and a real-world airline case study confirm that our approach not only outperforms classical RM heuristics in overall revenue but also significantly mitigates the buy-down effect by aligning fare offerings with evolving customer choices and competitive actions.

SA14

Building A Level 3 A316

Platform Operations

Job Market Showcase

Job Market Showcase

Chair: Emin Ozyoruk, University of Chicago, Chicago, United States

1 - Incentive Design on Digital Platforms: from Principal–Agent Theory to Livestreaming Commerce**Hang Jiang, National University of Singapore, Singapore, Singapore, Chen Jin, Luyi Yang**

This talk presents two studies bridging principal-agent (PA) theory and platform operations, focusing on livestreaming commerce.

The first paper revisits the classical first-order approach (FOA) for simplifying PA problems. We show the most celebrated Mirrlees-Rogerson condition validating FOA does not generalize to risk-neutral agents with limited liability. Motivated by our finding that quota-bonus contracts are optimal for the relaxed problem, we propose FOA validity specifically for quota-bonus contracts (FOAVQB). We provide general, economically intuitive, and easily verifiable conditions ensuring FOAVQB, thus guaranteeing quota-bonus contracts’ optimality.

The second paper analyzes optimal pricing, refund policies, and influencer contracts in livestreaming commerce under moral hazard. We demonstrate that the optimal refund policy is either full-refund or no-refund policies, contingent on product value. High-value products benefit from full refunds, which incentivizes more customer purchases and thus increases net sales. For low-value products, thin margins make no-refund policies desirable, as they ensure committed purchases and reduce inefficient production and return costs. Under no-refund policy, practically-used commission contracts (based on either gross or net sales) always achieve the theoretical global optimum. Under full-refund policies, however, gross and net sales commissions expose agents to different levels of uncertainty, making their effectiveness dependent on the agent’s risk aversion. Additionally, influencer misconduct (brushing) complicates gross-sales-based contracts under full refund policy. We then characterize optimal brushing-proof refund policies and contracts. Our findings provide strategic guidance for firms navigating livestreaming platforms’ nuanced operational challenges.

2 - Seller Inventory Competition Under Platform’s Algorithmic Demand Allocation**Shengyu Cao, Rotman School of Management, Toronto, ON, Canada, Ming Hu**

We explore how platform allocation rules influence seller behavior and shape inventory-based competition within online marketplaces. Through a detailed analysis of multiple sellers offering identical products on a platform, we investigate the effects of different allocation rules on seller inventory levels and platform payoff in the equilibrium. Contrary to expectations, findings reveal that inventory-agnostic priority rules can mirror sole-sourcing decision-making scenarios, suggesting that these allocation methods do not influence the platform’s optimal payoff. In contrast, incorporating randomness in demand allocation intensifies competition among sellers and increases the platform’s payoff. Moreover, the introduction of inventory-aware allocation rules offers a novel approach to demand allocation and further promotes competitive inventory strategies among sellers. We extend our analysis in two extensions. First, we examine how different platform allocation rules influence sellers, highlighting the trade-offs between platform payoff, customer fill rates, and both inter-temporal and cross-sectional profit disparities. This provides an alternative perspective on the design of allocation mechanisms. Second, we investigate a scenario where sellers must learn the allocation rules through reinforcement learning algorithms, which can lead to algorithmic collusion. We propose mitigation strategies and demonstrate how the discussed allocation rules can be adapted to sustain competitive marketplace in a dynamic environment.

3 - Risk and Management of Talent Drain: a Mixed Methods Approach

Nuo Yuan, Boston University, Boston, MA, United States, Mohamed Zaki Balboula, Yuze Li

Talent drain matters to firms as it disrupts operational stability, exacerbating financial performance variability to the detriment of shareholder value. Therefore, a precise understanding and management of imminent talent drain risk is essential to investors and firms themselves. Online crowd-sourcing platforms such as Glassdoor provide a centralized place for employees to reflect on their workplace experience (i.e., Glassdoor employee reviews) at their respective employers, providing investors and firms themselves with an alternative information intermediary to gauge talent drain risk. As such, we 1) empirically assess the extent to which investors detect and price in such risk by exploiting the timing a firm receives its first Glassdoor review (i.e., Glassdoor coverage initiation) as an exogenous shock and 2) employ an interpretable ML paradigm to identify operational "hotspots" most associated with a future talent drain event and propose a text-based future talent drain prediction algorithm. We contribute to the social media as information source for financial market literature by disputing prior finding via our finding that Glassdoor coverage initiation incurs an instantaneous spike in a firm's talent drain risk. We also contribute to the talent recruitment and retention literature by identifying salient operational drivers/remedies of future talent drain events and demonstrating the incremental predictive power of Glassdoor employee review text in predicting future talent drain events.

4 - Pricing and Capacity Decisions in Platform Competition with Network Externalities

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

The structure of network externalities influences platform competition and can determine whether a two-sided market is winner-takes-all or highly contestable. Yet, because of analytical challenges, relatively little is known about these structures in different markets. This paper investigates how the structure of externalities impacts the performance of platforms and competitive outcomes. We propose a model of platform competition in ride-hailing, identifying the structure of externalities that arise from congestion. After finding representations of congestion, which serve as micro-foundations for platform competition, we provide general conditions for the existence and uniqueness of equilibrium. Our results demonstrate that ride-hailing platforms benefit from economies of scale, becoming more efficient as they attract more riders and drivers. This could lead to winner-takes-all scenarios in markets with low demand or attractive outside options. However, service variability arising from differing distances between riders and drivers introduces differentiation, enabling multiple platforms to coexist in suitable markets. In such cases, entrants can leverage service variability to enter, potentially driving profits to zero and making the market highly contestable. The results challenge conventional wisdom, suggesting that having a large network may no longer guarantee survival and that platforms need credible commitment mechanisms to retain users. We discuss the reasons behind, shedding light on potential strategies that platform managers can follow while avoiding the ones that can inadvertently harm competition.

SA15

Building A Level 4 A401

Operational Optimization for Delivery Logistics in Urban Air Mobility and Commercial Aviation

Invited Session

Air Transportation Section

Chair: Shangqing Cao, University of California, Berkeley, Emeryville, CA, 94608, United States

Co-Chair: Shangqing Cao, University of California, Berkeley, Emeryville, CA, 94608, United States

1 - Map Partitioning for Same-Day Delivery Logistics System

Jianhua Liang, University of Texas at Austin, Austin, TX, United States

Same-Day Delivery (SDD) is a critical component of the logistics and transportation ecosystem, which garnered significant interest in the context of advanced air mobility (AAM) enabled E-commerce. Current research on SDD primarily addresses challenges in operational dispatch, leaving system design considerations largely unexplored. Previous studies on SDD system design usually considered a specific case with assumptions such as vehicle capacity and depot locations, which limit the implications of the models. These constraints also prohibit the exploitation of SSD with a single vehicle. In this work, we introduce a same-day delivery framework that incorporates vehicle capacity constraints and time guarantees, confirming a necessary schedule structure and bridging time and area constraints to ensure profitability. Tactically, we address the problem of what schedule and zone a vehicle should be responsible for to maximize the number of deliveries. We extend and generalize the existing same-day delivery model by including the vehicle's capacity constraint, as well as providing a set of comprehensive proofs for the necessity of optimality for the optimization problem. Besides, we also provide a straightforward implementation to convert the numerical optimization problem to one for maximizing profit. Finally, we provide an example demonstrating how system designers can achieve profitability while adhering to time guarantee SDD standards. This research helps reducing operational costs, improving delivery efficiency, while maximizing profitability. Further advancements based on this research, such as scaling to complex scenarios and optimal region partitioning, will further enhance the usage and enable real-world industry applications.

2 - From Voice to Safety: Language AI Powered Pilot-ATC Communication Understanding for Airport Surface Movement Collision Risk Assessment

John-Paul Clarke, University of Texas at Austin, Austin, TX, United States

Surface movement collision risk models are essential for airport safety, providing early warnings for near-miss incidents and reducing the risk of accidents. However, most existing models were developed decades ago and have not fully integrated recent advances in machine intelligence. To address this gap, we propose a language AI-based framework that enhances current Airport Surface Surveillance Capability (ASSC) by incorporating voice communication understanding for collision risk assessment.

Our framework has two components: (a) rule-enhanced Named Entity Recognition (NER) and (b) surface collision risk modeling. The NER module processes pilot-Air Traffic Controller (ATCo) voice transcripts to generate structured information tables, which inform taxi plans and collision risk calculations. We created and annotated a dataset from open-source recordings and safety reports, supplementing it with heuristic rules from FAA Order JO 7110.65W and FAA Order JO 7340.2N. Building on this, we developed the ATC Rule-Enhanced NER method, a hybrid model integrating heuristic rules into both training and inference, which outperforms baseline embedding models.

For risk modeling, we adopt a node-link airport layout graph (from NASA FACET) and model aircraft taxi speed as a log-normal distribution to estimate taxi times. We then propose a spatiotemporal formulation for collision probability when two aircraft approach conflict nodes. Case studies—including the January 2024 Haneda runway accident and September 2024 KATL taxiway collision—demonstrate that our model can estimate collision probabilities in real time. By combining voice communication understanding with surface movement analysis, our approach enables proactive measures to mitigate incursions and enhance airport safety.

3 - Scheduling and Routing for Multi-Modal Last-Mile Delivery Under Multiple Uncertainties

Minghao Chen, Columbia University, New York, NY, United States, Ang Li, Max Li

This research introduces a novel optimization framework for last-mile delivery using a tri-modal system that integrates trucks, drones, and tricycles. The proposed Stochastic Heterogeneous Vehicle Scheduling and Routing Problem (SHVSRP) addresses the operational complexities and synchronization risks arising from uncertain urban conditions such as traffic congestion and variable weather. We formulate SHVSRP as a mixed-integer program with dual objectives: minimizing expected delivery time and synchronization failures across modes. To solve this problem efficiently, especially at scale, we develop a two-phase parallel heuristic combining capacitated routing and time-aware scheduling. Simulation results across various traffic and wind conditions demonstrate that the tri-modal system significantly outperforms traditional truck-only and truck-drone models. Notably, it reduces makespan and failure rates, particularly under high uncertainty, thereby enhancing reliability and efficiency. Our study offers operational insights on optimal fleet composition and dynamic route planning, paving the way for resilient, scalable, and sustainable urban logistics.

4 - The UAM Fleet Scheduling Problem with Non-Linear Charging Time

Shangqing Cao, University of California, Berkeley, Emeryville, CA, United States

To better assess the economic feasibility of a UAM service, we introduce a vehicle-routing formulation of the Urban Air Mobility (UAM) fleet scheduling problem. The proposed model, electric urban air mobility vehicle routing problem with non-linear battery charging time (eUAMVRP-NL), studies the optimal flight scheduling and charging policy for a given fleet of eVTOL vehicles to maximize the operating revenue with the assumption of a non-linear charging model for the aircraft batteries. We design a Cluster-First Route-Second (CFRS) solution heuristic that first decomposes the k-vehicle problem into k number of 1-vehicle problems before optimizing the flight scheduling and charging decisions for each aircraft in the fleet. We validate our CFRS algorithm on a small, two-vertiport instance in which an optimal solution can be obtained. We found a loss of optimality less than 2%. Finally, we demonstrate the use of eUAMVRP-NL in obtaining meaningful metrics for the estimation of capital cost and operating cost by considering the airport access use case of UAM at Los Angeles International Airport (LAX).

5 - UCAirSim: Designing and Evaluating Cross-Campus UAM Network

Andrew Park, University of California, Berkeley, Berkeley, CA, United States

Urban Air Mobility (UAM) is emerging as a transformative mode of intra-regional transportation, enabled by electric vertical takeoff and landing (eVTOL) aircraft. This study presents UCAirSim, a discrete-event simulation framework designed to evaluate the performance of an inter-campus air taxi network connecting CITRIS campuses and major research hubs in Northern California. The framework integrates a Campus Interconnectivity Simulation (CIS)—a macroscopic, event-driven model of fleet scheduling, demand balancing, and energy dynamics—with a Guidance, Navigation, and Control (GNC) simulator that validates flight trajectories, aeropropulsive performance, and noise impacts using high-fidelity NASA eVTOL models.

The analysis evaluates system capacity, fleet sizing, and performance under fluctuating demand and varying flight paths. Results show that a fleet of 31 aircraft could serve approximately 400 passengers per day under nominal conditions, with scalability up to 1,600 passengers under optimized scheduling. Average wait times were reduced by incorporating repositioning strategies, while nonlinear battery charging dynamics constrained operational efficiency. Environmental and stakeholder considerations were also addressed, including vertiport siting, multimodal connectivity, and static noise policies. By combining network-level scheduling with vehicle-level dynamics, UCAirSim provides a scalable platform for analyzing operational strategies, environmental impacts, and future optimization approaches. This dual-layer simulation offers decision support for fleet operators, policymakers, and researchers in planning sustainable and efficient UAM systems for intraregional connectivity.

SA16

Building A Level 4 A402

AI: From Prototype to Production

Invited Session

The Practice Section of Informs

Chair: Ram Kumar Nimmakayala, WGU, Plano, TX, 75024, United States

1 - Bridging the Gap: Product-Led Execution for Enterprise AI Delivery

Ram Kumar Nimmakayala, WGU, Plano, TX, United States

Bridging the Gap: Product-Led Execution for Enterprise AI Delivery Despite rapid advancements in AI capabilities, many enterprise initiatives struggle to translate prototypes into scalable, value-generating solutions. This session focuses on a product-led execution framework that bridges the gap between experimentation and enterprise-grade AI delivery. It highlights how applying core product management principles—such as iterative validation, stakeholder alignment, user-centric design, and outcome-driven roadmaps—can transform fragmented AI efforts into cohesive, production-ready solutions. The discussion will explore real-world case studies, execution challenges in cross-functional environments, and techniques for aligning data science, engineering, and business teams around shared goals. Attendees will walk away with a pragmatic blueprint for operationalizing AI at scale, accelerating time-to-value, and embedding accountability throughout the AI lifecycle.

2 - Intelligent Automation of Data Engineering Workflows with AI

Manohar Sai Jasti, Workday, Cumming, GA, United States

Most data teams spend days turning a helpful query into a production-grade pipeline—refactoring code into dbt models, adding data-quality tests, writing documentation, and scheduling the flow in Airflow. These manual steps slow insights and tie up scarce engineering time.

This talk shows how large language models remove much of that toil. An LLM can

- Generate modular dbt models (staging, intermediate, mart) from draft SQL
- Insert freshness, null-check, and uniqueness tests
- Write concise model documentation and lineage notes
- Build a ready-to-run Airflow DAG

Attendees will see how automating these routine tasks speeds delivery of reliable data products and frees engineers to focus on higher-value analysis.

3 - Running Away from Demo Hell: Shipping AI Beyond the Prototype with Eval-Driven Development

Albert Lie, Forward Labs, San Francisco, CA, United States

Many AI projects shine in demos but fail in production, trapped in what we call "Demo Hell." The transition from a promising prototype to a reliable, real-world AI system is riddled with unseen pitfalls, from brittle prompts to unpredictable model outputs. In this talk, we'll explore Eval-Driven Development (EDD) as the key to escaping Demo Hell and building AI that actually works beyond controlled environments. Using a few open-source frameworks for evaluating LLM applications, we'll dive into: Why traditional AI development often breaks down in real-world use cases How Eval-Driven Development ensures continuous, measurable improvement Setting up automated evaluations to diagnose failures before they reach users Case studies of teams that successfully shipped AI from prototype to production using EDD

4 - A Framework for Generative AI Operation Evaluation

Cigil Achenkunju, Northern Illinois University, Boston, MA, United States

How do we know if the investment in generative artificial intelligence (GenAI) is moving the needle? It is a question heard almost daily across finance, healthcare, and retail sectors. And honestly, it is the right question to ask. On top of that, should we continue to invest in AI at the same rate or optimize? How can a stakeholder show that solution usage has a positive or negative impact on their operation?

Survey results indicate that up to 85% of AI initiatives eventually fail to deliver their promises. Organizations using gen AI want to understand the impact of such

solutions clearly. Can you blame them? So, let's define a strategic decision-making framework that broadly answers these business questions in an operational setting that balances the benefits of business value and AI integration.

An analytical framework for operations measurement 2S/2E: Think of it as four pillars used together to tell a complete story of your gen AI-enabled operation's health: Each pillar reveals a different facet of your performance, and I'll show you exactly how to measure them. These pillars offer valuable insights to measure your operations. What makes this framework powerful is its systematic approach and adaptability.

SA17

Building A Level 4 A403

Labor Planning and UTR Process Optimization

Invited Session

Practice Track

Chair: Elcin Cetinkaya, Amazon.com, Seattle, WA, United States

1 - Optimizing the Allocation of Outbound Work to Constrained Resources in Amazon Fulfillment Centers

Kevin Bunn, Amazon, Bellevue, WA, United States

In Amazon Fulfillment Centers, there are multiple options for how to prepare a shipment to be sent to a customer. Each of these processing options is called an outbound process path. Every process path has a planned throughput, which can be used to determine the capacity available for how many shipments can be processed on that process path in a given amount of time. In this presentation, we will discuss an optimization-based planning model which is used to plan the allocation of shipments to each process path, given the set of process path options for those shipments. We will discuss the main objectives of this optimization model, and the factors which led to the decision to use an optimization model to solve this problem.

2 - Prior Day Labor Planning in Amazon Fulfillment Center

Yingqi Zhang, Amazon, Bellevue, WA, United States, Andy Johnson

This paper presents an innovative optimization approach to Prior Day Planning (PDP) in large-scale operations management. While labor planning typically spans multiple weeks, we specifically address the critical 24-hour window before execution, where site operators must finalize staffing decisions to meet committed volume targets. The existing PDP process, traditionally conducted through manual collaboration among department leaders, lacks systematic technological support for optimal decision-making. To address this gap, we introduce a mixed-integer programming (MIP) model that generates optimal labor allocation plans at granular working period and process levels. The model systematically evaluates cross-departmental labor sharing opportunities and staffing requirements, providing data-driven recommendations for resource allocation. Our solution integrates four key operational services into a streamlined planning framework, designed to enhance operational efficiency, optimize workforce utilization, and maintain service quality standards while responding to dynamic market conditions. This automated approach represents a significant advancement from manual planning methods, offering a more robust and systematic solution to daily labor optimization challenges.

3 - Optimizing Order Picking for Amazon Robotics Fulfillment Centers

Ignacio Neira, Amazon.com, Bellevue, WA, United States

Fulfilling orders at the scale and speed that Amazon does is a challenging task that requires optimized processes throughout an order's life cycle. Amazon Robotics (AR) Fulfillment Centers (FC) have proved to be more efficient and ergonomics-friendly than its traditional counterparts, empowered by the robotic technology that enables stationary and continuous picking from inventory pods that are brought to pickers by robotic drives. AR FCs face different challenges than traditional FCs and requires innovative ways to solve for a pick execution plan quickly while accounting for associate safety. Today's pick planning in AR FCs is driven by a heuristic that can quickly create short-term pick execution plans to meet upstream targets; however, its greedy nature has shown to be very limited in leveraging all the data and optimizing for multiple objectives. In this talk, we present a new optimization-based, scalable, and flexible framework for AR pick planning and execution that coherently optimizes for upstream targets adherence, pick productivity and associate safety.

4 - Robust and Interpretable Outbound Resource Plans

AARON HERMAN, Amazon, Seattle, WA, United States

A high quality, interpretable resource plan is a crucial component for a successful shift in Amazon Fulfillment Centers (FCs). A plan is high quality if it meets key objectives such as processing shipments in time and maximizing labor efficiency, while being robust to the uncertainty which is an inherent feature of warehouse operations. It must also be interpretable for managers in warehouses who are responsible for executing the decisions of the plan. In this talk we will describe components of a modeling strategy that meets the goals of being robust and interpretable.

5 - Computer Vision Integrated Warehouse Emulation for Bulk Package Flow

Sumant Joshi, Amazon.com, Bellevue, WA, United States, Chin-Yuan Tseng

Material flow simulation and emulation serve as cornerstone technologies in automated warehouse design, providing environment for digital twin development and control logic validation through production system integration. With recent years rapid growth in computation power and graphic processing algorithms, vision systems that capture real-time images and videos are massively deployed in robotic pick/place and bulk flow control applications. One of the key areas where Material flow simulation and emulation gap exists, is modeling vision systems to estimate volumetric information for bulk flow applications. In most simulation software, the image captured by vision camera cannot be analyzed locally for which simplified sensors are used which may not represent real system behavior. To enhance controls emulation, real-time vision data must be processed by AI/ML algorithms, with resulting insights driving decision-making in PLCs or industrial PCs. This research presents a novel integration framework between Emulate3D and computer vision algorithms, enabling real-time bidirectional communication for vision-based control. Our approach seamlessly transfers actual visual data from warehouse to Emulate3D using advanced image processing. Based on actual image processing, similar package setup is generated in Emulate3D for further emulation based on PLC logic. This integrated approach significantly enhances emulation fidelity and control system reliability as well as providing a platform for computer vision development.

SA18

Building A Level 4 A404

Intelligent Transportation Systems and Cybersecurity

Invited Session

TSL: Intelligent Transportation Systems

Chair: Mohamadhossein Noruzoliaee, University of Texas, Rio Grande Valley, Edinburg, TX, United States

1 - Digital Twin-Based Driver Risk-Aware Intelligent Mobility Analytics for Urban Transportation Management

Zilin Bian, Rochester Institute of Technology, Rochester, NY, United States

Traditional mobility management strategies emphasize macro-level mobility oversight from traffic-sensing infrastructures, often overlooking safety risks that directly affect road users. To address this, we propose a Digital Twin-based Driver Risk-Aware Intelligent Mobility Analytics (DT-DIMA) system. The DT-DIMA system integrates real-time traffic information from pan-tilt-cameras (PTCs), synchronizes this data into a digital twin to accurately replicate the physical world, and predicts network-wide mobility and safety risks in real time. The system's innovation lies in its integration of spatial-temporal modeling, simulation, and online control modules. Tested and evaluated under normal traffic conditions and incidental situations (e.g., unexpected accidents, pre-planned work zones) in a simulated testbed in Brooklyn, New York, DT-DIMA demonstrated mean absolute percentage errors (MAPEs) ranging from 8.40% to 15.11% in estimating network-level traffic volume and MAPEs from 0.85% to 12.97% in network-level safety risk prediction. In addition, the highly accurate safety risk prediction enables PTCs to preemptively monitor road segments

with high driving risks before incidents take place. Such proactive PTC surveillance creates around a 5-minute lead time in capturing traffic incidents. The DT-DIMA system enables transportation managers to understand mobility not only in terms of traffic patterns but also driver-experienced safety risks, allowing for proactive resource allocation in response to various traffic situations. To the authors' best knowledge, DT-DIMA is the first urban mobility management system that considers both mobility and safety risks based on digital twin architecture.

2 - Certifiable Hybrid Cyber Attack and Defense on Learning-Based Traffic Signal Control

Mohamadhossein Noruzoliaee, University of Texas Rio Grande Valley, Edinburg, TX, United States, Mark Hernandez, Fatemeh Nazari

Learning-based control methods, such as deep reinforcement learning (DRL), have gained traction in intelligent transportation systems (ITS) control due to their computational efficiency in dealing with the scale and dynamics of ITS. Despite growing popularity, DRL has been recently found to be vulnerable to cyber-attacks, rendering its deployment challenging in safety-critical ITS unless its adversarial vulnerabilities are mitigated. This research aims to robustify DRL-based traffic signal control (i.e., victim) against "hybrid" cyber-attacks on both the input state observations (e.g., traffic queue length at an intersection) and output actions (e.g., traffic signal phase selection) of the victim.

3 - Distributed Versus Centralized Optimization Frameworks for Maximum-Stability Dispatch of Shared Autonomous Electric Vehicles

Jake Robbennolt, University of Texas at Austin, Austin, TX, United States, Stephen Boyles, Javad Mohammadi

As shared autonomous electric vehicle (SAEV) fleets scale, centralized dispatch strategies face increasing computational burdens and communication bottlenecks. This is particularly problematic when optimal dispatch strategies incorporate time-varying factors including passenger demand patterns, vehicle battery states, electricity pricing fluctuations, and charging station availability which necessitate model predictive control frameworks to ensure future conditions are accounted for. With these concerns in mind, this research investigates the fundamental trade-offs between centralized and distributed optimization approaches for SAEV fleet management while maintaining maximum-stability guarantees—ensuring passenger queues remain bounded for any feasible demand pattern. We propose a distributed consensus+innovations method where individual vehicles and charging stations operate as computational agents in a multi-agent network, collaboratively solving the dispatch optimization problem without requiring a central controller. The proposed method maintains theoretical maximum-stability guarantees—ensuring passenger queues remain bounded for any feasible demand pattern—while distributing computational load across the network. Through rigorous mathematical analysis, we establish connections between communication burden, network topology, and system-level performance, identifying conditions where distributed approaches outperform centralized alternatives. Simulation experiments demonstrate that our distributed approach scales better with fleet size while maintaining or improving performance metrics compared to centralized dispatch policies.

4 - From Openlka to Lkalert: An End-to-End Pipeline for Real-World Lane-Keeping Safety

Yuhang Wang, Univeristy of South Florida, Tampa, FL, United States, Hao Zhou, Abdulaziz Alhuraish

Lane-Keeping Assist (LKA) is now standard on most market vehicles and has measurably reduced lane-departure crashes, yet production systems still disengage without warning in adverse conditions, forcing instantaneous driver take-overs and leaving researchers without holistic evidence of failure modes. We address these gaps with an end-to-end pipeline that unites open data, failure analytics, and proactive warning system. The OpenLKA dataset, released by us, is the first public corpus devoted to LKA evaluation, offering 400 h of time-aligned CAN telemetry, front-view video, perception outputs, and vision-language annotations from 62 vehicle models across varied weather, illumination, and geometry. Using this resource, we deliver the first comprehensive field study of LKA behaviour, catalogue high-risk scenarios, and verify their impact via statistic analysis; the derived features power a supervised model that assigns an infrastructure safety index to road segments, highlighting maintenance priorities. To restore reaction time for drivers, we curate OpenLKA-Alert, a 55 k-frame expert-labelled subset of imminent disengagements, and develop LKAAlert, a multimodal early-warning system built on Qwen2.5-VL-7B with surrogate-feature guidance and LoRA fine-tuning. LKAAlert forecasts LKA failures 1-3 s ahead with 69.8 % accuracy, operates at real-time 2 Hz on embedded hardware, and produces natural-language rationales scoring 71.7 ROUGE-L, demonstrating strong predictive power and interpretability in tandem. Collectively, OpenLKA, its analytical toolbox, and LKAAlert constitute the first integrated benchmark, diagnostic framework, and mitigation suite for real-world lane-keeping safety.

5 - Usage-Based Insurance in the Era of Connected Vehicles: Competition and Data Differentiation

Sen Yan, New York University, New York City, NY, United States, Daniel Vignon

Traditional auto insurance models have long relied on socio-demographic proxies to assess risk, a method that inadequately captures the true determinants of accident risk-driving behavior. With telematics technology, insurers can capture real-time driving data to offer usage-based insurance (UBI). In addition, direct access to extensive vehicle and driving behavior data incentivizes automakers to enter the market. Since the latter potentially possess better information on drivers than insurers, one rightfully wonders about (a) the extent to which this difference in consumer data affects the market and (b) policies to remediate potential distortions in this context. To answer these questions, we model competition between insurers whose costs of collecting consumer behavior data differ. We explore how this difference affects not only the equilibrium distribution of customers across firms, but also the overall welfare. We find that, though monitoring itself is welfare improving in the case of a monopoly, it may reduce welfare in the case of a duopoly. While requiring data-sharing might be welfare improving, it is only so when (a) investment in data collection technology is low or (b) when sharing is priced. These findings provide key insights into the welfare benefits of UBI, the competitive dynamics between traditional insurers and automakers, and policy implications regarding data sharing in the connected vehicle era.

SA19

Building A Level 4 A405

Applied Probability in Algorithmic and Operational Networks I

Invited Session

Applied Probability Society

Chair: Yeganeh Alimohammadi, USC, Los Angeles, CA, United States

Co-Chair: Sophie Yu, The Wharton School of the University of Pennsylvania, Philadelphia, PA, United States

1 - Endogenous Entry in Networked Markets with Production and Edge Capacity Constraints

Cheng Guo, Clemson University, Clemson, SC, United States, Ozan Candogan, Jiayi Wang

Motivated by electricity markets, we investigate the key drivers of firm entry and investment in multi-regional markets with elastic demand, incorporating capacity constraints for both firms and transportation. These constraints introduce externalities due to the coupling of decisions across firms and regions. To address such complication, we develop a novel convex optimization model for networked Cournot competition, which is substantially more tractable than the commonly-used mixed complementarity problem formulations. This model enables us to prove the uniqueness of the equilibrium, analyze how factors such as entry locations and tariffs shape investment decisions, and carry out a case study on a realistic electricity markets dataset. We find that a well-chosen firm entry location not only benefits the firm's own profit, but may also bring positive externalities for incumbent firms and consumers. Additionally, we show that import tariffs may, counterintuitively, encourage the expansion of more efficient foreign firms. Our results offer both novel analytical frameworks and scalable computational tools for practitioners, guiding efficient investment of capacity and infrastructure in networked markets.

2 - Optimal Mechanisms for Demand Response: An Indifference Set Approach

mohammad mehrabi, Stanford GSB, Stanford, CA, United States, Omer Karaduman, Stefan Wager

The time at which renewable (e.g., solar or wind) energy resources produce electricity cannot generally be controlled. In many settings, consumers have some flexibility in their energy consumption needs, and there is growing interest in demand-response programs that leverage this flexibility to shift energy consumption to better match renewable production—thus enabling more efficient utilization of these resources.

We study optimal demand response in a model where consumers operate home energy management systems (HEMS) that can compute the “indifference set” of energy-consumption profiles meeting pre-specified consumer objectives, receive demand-response signals from the grid, and control consumer devices within the indifference set.

For example, if a consumer asks for the indoor temperature to remain between certain upper and lower bounds, a HEMS could time the use of air conditioning or heating to align with high renewable production when possible.

Here, we show that while price-based mechanisms do not generally achieve optimal demand response—i.e., dynamic pricing cannot induce HEMS to choose optimal demand-consumption profiles within the available indifference sets—pricing is asymptotically optimal in a mean-field limit with a growing number of consumers. Furthermore, we show that large-sample optimal prices can be efficiently derived via an algorithm that only requires querying HEMS about their planned consumption schedules under different prices. Finally, we study learning the optimal price passively (when frequent queries to HEMS are not possible) as an online learning problem over a series of days, using only the day's forecasts—i.e., weather, temperature, and weekday conditions—as context, and by observing only consumer consumption profiles.

3 - Mallows Model with Learned Distance Metrics: Sampling and Maximum Likelihood Estimation

Yeganeh Alimohammadi, USC, Los Angeles, CA, United States

The Mallows model is a widely-used probabilistic framework for learning from ranking data, with applications ranging from recommendation systems and voting to aligning language models with human preferences [1]. Under this model, observed rankings are noisy perturbations of a central ranking σ , with likelihood decaying exponentially in distance from σ , i.e., $P(\pi) \propto \exp(-\beta \cdot d(\pi, \sigma))$, where $\beta > 0$ controls dispersion and d is a distance function.

Existing methods mainly focus on fixed distances (such as Kendall's τ distance), with no principled approach to learning the distance metric directly from data. In practice, however, rankings naturally vary by context; for instance, in some sports we regularly see long-range swaps (a low-rank team beating a high-rank one), while in others such events are rare.

Motivated by this, we propose a generalization of Mallows model that learns the distance metric directly from data. Specifically, we focus on L_α distances: $d_\alpha(\pi, \sigma) = \sum_{i=1}^n |\pi(i) - \sigma(i)|^\alpha$.

For any $\alpha \geq 1$ and $\beta > 0$, we develop a Fully Polynomial-Time Approximation Scheme (FPTAS) to efficiently generate samples that are ϵ -close (in total variation distance) to the true distribution. Even in the special cases of L_1 and L_2 , this generalizes prior results that required vanishing dispersion ($\beta \rightarrow 0$).

Using this sampling algorithm, we propose an efficient Maximum Likelihood Estimation (MLE) algorithm that jointly estimates the central ranking, the dispersion parameter, and the optimal distance metric. We prove strong consistency results for our estimators, and we validate our approach empirically using datasets from sports rankings.

4 - Effectiveness of Interventions in Epidemic Spread

Karissa Huang, University of California, Berkeley, Berkeley, CA, United States

As the world grows increasingly connected, infectious disease transmission and outbreaks become a pressing global concern for public health officials and policymakers. While policy interventions to contain and prevent the spread of disease have been proposed and implemented, there has been little rigorous quantitative analysis of the effectiveness of such interventions. In this talk, we discuss the susceptible-infected-recovered (SIR) infection process on a dynamic network model that models two communities with travel between them. In particular, we consider two Erdos-Renyi graphs where edges are dynamically changing in each graph based on node travel between the graphs. We characterize the time evolution of the outbreaks in both communities and pin down the time for when the infection first reaches the second community. Finally, we analyze two interventions—social distancing and travel bans—and show that while social distancing is effective at reducing the burden of the disease in the second community, travel bans are not.

5 - TBD

Anirudh Sridhar, MIT, Cambridge, MA, United States

TBD

SA20

Building A Level 4 A406

Machine Learning and Urban Network Analytics

Invited Session

Telecommunications and Network Analytics

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

1 - Data-Driven Dynamic Modeling of Chaotic Systems Using Local Maximum Entropy Expansion

Nazia Raza, Lehigh University, Bethlehem, PA, United States, Farrah Moazeni

Learning the dynamics of real-world systems without prior knowledge remains a fundamental challenge across engineering disciplines. Traditional identification techniques often rely on strong assumptions or detailed physical models, limiting their applicability in chaotic, complex, or partially observed environments. Meanwhile, the rise of sensor networks and IoT infrastructure has led to abundant high-resolution time-series data, much of which remains underutilized. This work introduces a novel data-driven framework for chaotic system identification based on Local Maximum Entropy (LMaxEnt) modeling, which requires no structural assumptions and integrates principles from information theory and polynomial interpolation. The method learns time-resolved system derivatives using locally weighted basis functions, forming a flexible nonparametric surrogate model. To control locality and smoothness, we incorporate three prior kernels (Gaussian, Laplacian, and Cauchy) and introduce a tunable locality parameter that governs each basis node's influence, enabling accurate approximation of both smooth and stiff dynamics. We validate the proposed framework on the classical quadruple tank process (QTPS) and a modified QTPS with additional coupling, against the chaotic Lorenz system as a benchmark. Across all studied systems, LMaxEnt demonstrates high-fidelity reconstruction of system trajectories and derivatives. The approach is experimentally validated using real-time data from a lab-scale QTPS and a water distribution network testbed, highlighting its robustness to noise and practical deployment. This study presents a generalizable, interpretable, and assumption-free pathway for modeling complex, chaotic dynamical systems, with promising applications in control, monitoring, and digital twin development for cyber-physical systems.

2 - Towards a Generalized Network Optimization Model for Community Resilience Decision Support

Kenneth Harrison, National Institute of Standards and Technology (NIST), Gaithersburg, MD, United States, Tasnim Faiz, William Hughes

Natural disasters, like floods, tornadoes, earthquakes, and wildfires, wreak havoc on a community's interdependent physical, social, and economic systems, or networks. These natural hazards present loadings (e.g., winds, inundation, ground acceleration) that cause network components to fail, in turn triggering outages impairing the social functions of the community (e.g., housing, businesses, and healthcare); importantly, a protracted period of recovery typically follows. NIST has developed several network optimization models to support community resilience decision-making. Recent advances include modeling the voluntary uptake of mitigation actions (e.g., new building retrofits) under incentive programs, addressing the extended implementation period of mitigation decisions, modeling access to healthcare services, and giving decision-makers tools for addressing risk. This talk describes work towards a generalization of the model for application to multiple communities confronting a range of hazards. Various complicating issues are addressed, including network component failure modeling differences across hazards, aligning with the community's other planning objectives, and satisfying multiple stakeholders. Case studies of communities facing various hazards demonstrate this work towards a common decision support framework for community resilience planning.

3 - Restoring Complex Flow Networks Using Deep Reinforcement Learning

Amir Shahlaee, Department of Civil Engineering, University of Texas at Arlington, Arlington, TX, United States, Xudong Fan, Wang Yu, Jin-Zhu Yu, Jianxi Gao

Ensuring the rapid and efficient recovery of large-scale complex flow networks following disruptions is critical to the resilience of those networks. While optimization models such as Mixed Integer Programming (MIP) offer accurate solutions, their computational intractability on large-scale, flow-constrained networks hinders practical application. This paper presents a Deep Reinforcement Learning (DRL) framework, guided by MIP during training, to learn optimal restoration sequences for damaged infrastructure. Our proposed MIP-guided Deep Q-Network (DQN) agent, equipped with a prioritized replay buffer, achieves fast and robust performance. The agent is trained on 60 small synthetic networks and evaluated on larger ones, demonstrating superior performance over traditional optimization approaches in both efficiency and solution quality. This work contributes a scalable, DRL-based methodology for enhancing the resilience of infrastructure systems subject to complex physical constraints such as flow conservation and traffic assignment.

SA21

Building A Level 4 A407

Advances in Learning-Enhanced Combinatorial Optimization for Operations

Invited Session

Transportation Science and Logistics (TSL)

Chair: Yu Yang, University of Florida, 401C, Weil Hall, Gainesville, FL, 32611-0001, United States

Co-Chair: Ricky You, The University of Florida, Bellevue, WA, United States

1 - Two-Stage Learning to Branch in Branch-Price-and-Cut Algorithms for Solving Vehicle Routing Problems Exactly

Yu Yang, University of Florida, Gainesville, FL, United States

Branching is one of the most important components in branch-price-and-cut (BPC) algorithms for solving vehicle routing problems (VRPs) exactly. However, learning to branch is much more challenging in BPC than in branch-and-cut algorithms that are used for solving general mixed integer programs. To address such challenges, we propose the first effective learning-to-branch framework in BPC algorithms, leading to a novel two-stage learning-based branching (2LBB) strategy. Moreover, we propose a novel theoretical model characterizing the fundamental trade-off between time spent making a single branching decision and the resulting branching quality. Through an extensive numerical study, we demonstrate that a dynamic version of the 2LBB, denoted by 2LBB-dy, achieves approximately 45% and 50% time reduction, respectively, compared to the state-of-the-art (SOTA) hand-crafted branching strategy in solving the capacitated vehicle routing problem (CVRP) and vehicle routing problem with time windows (VRPTW). In addition, our RouteOpt (an exact VRP solver available at <https://github.com/Zhengzhong-You/RouteOpt>), when equipped with the 2LBB-dy, achieves a 47% time reduction compared to the SOTA VRPSolver from Pessoa et al. (2020) for the CVRP.

2 - Accelerating Column Generation for Vehicle Routing Problems Using Graph Neural Network-Predicted Dual-Optimal Inequalities

Ricky You, University of Florida, Gainesville, FL, United States, Bo Tang, Yu Yang

Column generation (CG) is widely used for solving combinatorial optimization problems due to strong dual bounds provided, but it often suffers from slow convergence caused by fluctuations in dual variables, especially early in the solution process. To stabilize CG, previous works introduced dual-optimal inequalities (DOIs) for structured problems like vector packing. However, vehicle routing problems (VRPs) lack the explicit structure required for deriving these inequalities analytically. We propose using graph neural networks (GNNs) to predict potential DOIs for customer pairs, effectively generating learned dual constraints. A graph-based reduction method further simplifies these constraints, preserving same dual feasible region while substantially reducing the formulation size. Computational experiments on standard VRP benchmarks demonstrate significant speedups; notably, our method reduces root-node computation time for a large-scale VRP instance (X-n1001-k43) from hours to minutes.

3 - Optimizing on-Demand Flexible Transit Services via Graph-Attention Multi-Agent Proximal Policy Optimization

Wen Zhang, The State University of New York at Buffalo, Buffalo, NY, United States, Jee Eun Kang

This study proposes a data-driven, real-time optimization framework for on-demand flexible transit services by integrating Graph Attention Networks (GAT) with Multi-Agent Proximal Policy Optimization (MAPPO). In contrast to traditional Vehicle Routing Problem with Pickup and Delivery (VRPPD) and Dial-A-Ride Problem (DARP) models, which assume static and pre-scheduled trips, our method models each vehicle as an autonomous agent operating within a Semi-Markov Decision Process. Tasks are dynamically allocated based on real-time system states and historical travel demand patterns. Vehicle states along with inter-vehicle dependencies are encoded via a shared graph-based critic, while decentralized agents select actions individually from a shared actor policy based on their local observations. The framework is trained end-to-end using Automated Fare Collection (AFC) data from Seoul with the explicit objective of minimizing fleet-wide total travel time and demonstrates significant improvements in operational efficiency. The proposed approach offers a scalable, adaptive solution for real-time transit operations.

4 - Deep-Learning Aided Crowdsourcing Network Design with External Delivery Vehicles

Baris Yildiz, Koc University, Istanbul, Turkey, Ramazan Cevik, Ali Sardag

Crowdsourcing has gained traction for its economic and environmental benefits, offering cost-effective deliveries while reducing traffic and emissions. However, the sporadic arrival of crowdsourcers limits reliability, especially on routes with heavy package flows. To address this, we propose a hybrid delivery model that integrates crowdsourcing with dedicated delivery vehicles along selected routes to improve system efficiency. Our goal is to identify which segments of the delivery network should be supported by external vehicles, determine the optimal number of vehicles, and design their routes. We formulate the problem on a time-extended network and solve it in two stages. In the first stage, we develop a Dynamic Programming (DP) algorithm that determines the optimal routing and compensation for crowdsourcers, given the locations of the supported routes. This problem is modeled as a Markov Decision Process (MDP), which we solve efficiently by leveraging the acyclic structure of the network. In the second stage, we generate multiple random network configurations and solve the MDP for each, creating a dataset of total crowdsourcing costs. A Feedforward Neural Network (FNN) is then trained to predict these costs based on network characteristics. This FNN is embedded into a Mixed Integer Programming (MIP) model to jointly optimize route selection and vehicle planning. We demonstrate that the FNN achieves competitive performance in cost prediction. Under mild conditions, it can be incorporated into a Mixed Integer Quadratic Programming (MIQP) model through appropriate transformations without compromising model performance.

5 - Regularized Langevin Dynamics for Combinatorial Optimization

Shengyu Feng, Carnegie Mellon University, Pittsburgh, PA, United States

This work proposes a simple yet effective sampling framework for CO. Our method builds on discrete Langevin dynamics (LD), an efficient gradient-guided sampling algorithm. However, we observed that directly applying discrete LD often leads to limited exploration. To overcome this limitation, we propose the *Regularized Langevin Dynamics (RLD)*, which enforces an expected distance between the sampled and current solutions, effectively avoiding local minima. We develop two CO solvers on top of RLD, one based on simulated annealing (SA) and the other one based on neural network (NN). Empirical results on three classical CO problems demonstrate that both of our methods can achieve comparable or better performance against the previous state-of-the-art (SOTA) SA and NN-based solvers. Our SA algorithm reduces the running time of the previous SOTA SA method by up to 80%, while achieving equal or superior performance. And our NN solvers achieve a comparable performance to the previous SOTA NN methods with only 5% training time due to its lightweight training objective. The extension of RLD also demonstrates its applicability to more general CO problems such as mixed integer programs or the ones with global constraints. In summary, RLD offers a promising framework for enhancing both traditional heuristics and NN models to solve CO problems.

SA22

Building A Level 4 A408

Frontiers in Data-Driven Resource Allocation and Inventory Management

Invited Session

Applied Probability Society

Chair: Sean Sinclair, Northwestern University, Evanston, IL, United States

1 - Sample Complexity of Policy Learning for Inventory Control with Censored Demand

Xiaoyu Fan, Stern School of Business, New York University, New York, NY, United States, Boxiao Chen, Zhengyuan Zhou

Traditional offline-learning in inventory control assumes access to uncensored demand samples, which are often not available, as the firm can only access sales data (i.e. censored demand data). In this paper, we study the classic newsvendor problem in an offline learning setting with censored demand. We design an offline algorithm that achieves the optimal sample complexity under both iid data and adaptive data. We also extend our results to the infinite-horizon setting.

2 - Online Learning for Lost-Sale Inventory Systems with Product Returns

Qi Luo, University of Iowa, Iowa City, IA, United States, Boxiao Chen, Stefanus Jasin, Mengxiao Zhang

Product returns pose a significant challenge to supply chain sustainability and efficiency, as evidenced by \$744 billion worth of returned merchandise in 2024. To address this, we propose an online learning algorithm for single-item inventory control that adapts to demand censoring and unknown return rates over a horizon of T periods. The algorithm sequentially updates the base-stock policy, achieving asymptotic optimality with a worst-case regret bound of $O(T^{0.5})$. Our approach addresses the technical challenges and practical importance of incorporating return behavior into learning-based inventory control policies.

3 - Online Allocation of Throughput-Constrained Resources Using Proxy Assignments

Chamsi Hssaine, University of Southern California, Marshall School of Business, Los Angeles, CA, United States, Kamessi Zhao, Daniel Freund

Effective load balancing is at the heart of many applications in operations. Frequently tackled via the balls-into-bins paradigm, seminal results established the power of two choices in load balancing: a limited amount of costly flexibility goes a long way to maintain approximately balanced loads throughout the decision-making horizon. This paper is motivated by the fact that balance across time is too stringent a requirement for some applications; rather, the only desideratum is approximate balance at the end of the horizon. Thus motivated, in this work we design "delayed-flexibility" algorithms for such settings. For the canonical balls-into-bins problem, we show that a simple policy that begins exerting flexibility toward the end of the time horizon (i.e., when $\Theta(\sqrt{T \log T})$ periods remain), suffices to achieve an approximately balanced load (i.e., a maximum load within $O(1)$ of the average load). Moreover, with just a small amount of adaptivity, a threshold policy achieves the same result, while only exerting flexibility in $O(\sqrt{T})$ periods, thus matching a natural lower bound. We leverage these results to study the design of opaque selling strategies in retail settings, which have recently been found to be a prime application of the power of two choices in balls-into-bins. For this problem, we prove that late-stage opaque selling strategies achieve the optimal trade-off between exerting costly flexibility (i.e., the revenue loss from offering the opaque product at a discount) and achieving strong inventory cost savings due to load balancing. We demonstrate the robustness of our insights via extensive numerical experiments.

4 - Managing Inventory Systems with Stochastic Lead Times and Impatient Customers

Jinzhi Bu, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Guangyu Li

In today's global supply chain, lead time uncertainty presents significant challenges for online retailers, especially in managing inventories while facing impatient customers. We address this by studying a periodic-review inventory system with stochastic lead times and partial backorders in an online retail environment. The optimal policy for this system is in general very complex. We propose simple heuristic replenishment policies and analyze their performances. In the full distributional-information setting, we establish the asymptotic optimality of these policies with large unit shortage penalty costs. In the data-driven setting, we develop online learning algorithms that achieve tight regret bounds.

5 - Retail on Wheels: A Time-Sensitive Travelling Salesman Problem

Zhuolun Dong, University of Texas at Austin, Austin, TX, United States

Rapid advancement in self-driving technology has brought mobile retail into the spotlight. Motivated by its tremendous potential, we present a model, theory, and insights into the operations of the business model, where a fleet of wheeled stores traverses a service region, selectively seeks locations in which to perch, and fulfills local customer demand. The goal of retailers is to balance between perching and migrating to maximize total revenue. To dispatch such stores, retailers need to consider the following time-sensitive factors: (i) How busy the service is, (ii) how soon customers come to the store, and (iii) how fast the product decays. We propose and analyze a model that incorporates these factors into the classical traveling salesman problem (TSP), which we call the "time-sensitive TSP" (TSTSP). Using the continuous approximation approach, we derive the optimal revenue as well as the upper and lower bounds in the asymptotic regime as candidate locations increase to infinity. The tightness of the bounds is validated. Our analytical and numerical analysis demonstrates that store mobility creates value by bringing proximity to customers and by taking advantage of flexible repositioning to reach high-profit areas. We further uncover the effects of these time-sensitive features on decision-making. Specifically, the scaling relationship between revenue and fleet size is affected by customer tardiness and busyness. Busyness determines stores' ability to expand their service, whereas tardiness affects how effectively stores can utilize the initial high-freshness period. Our findings demonstrate the potential of mobile retail to emerge as a new competitive retail form.

SA23

Building A Level 4 A409

Analytics for the Great Good

Invited Session

Minority Issues Forum

Chair: M Gabriela Sava, Bowling Green State University, Bowling Green, OH, United States

1 - Continuity of Hospital Care and Patient Outcomes: Evidence from Cabg Cases

Dawson Maddox, University of Tennessee, Knoxville, TN, United States, Bogdan Bichescu, Haileab Hilafu

We examine the relationship between continuity of hospital care (CoHC) and outcomes for patients undergoing coronary artery bypass grafting (CABG). Using linked inpatient and outpatient data, we define CoHC as receiving care from the same hospital across both settings. We find that greater CoHC is associated with shorter lengths of stay and lower readmission risk, though it has no significant effect on total charges. These relationships are moderated by physician utilization intensity and procedural focus.

2 - Bias in Predictive Models that use Electronic Health Record (EHR) Data

M Gabriela Sava, Bowling Green State University, Bowling Green, OH, United States, Azeez Kunle Akinbode, Jerrold May, Ronald Pirrallo

Methods for predictive modeling typically assume that all patient records are complete. However, EHR records may be incomplete, especially for groups of patients, such as for minorities and for those in lower socioeconomic groups who have only sporadic interaction with the healthcare system. We investigate, in the context of preventive screening, the impact of incomplete information on modeling, and the biases that might be introduced by predictive models created from, or applied to, incomplete patient records. From a clinical perspective, our research is of interest, because some best practice alerts need to be triggered even when only partial patient information is available.

3 - Identifying Illicit Drug Trafficking Patterns Using Geographical Weighted Regression: a Case Study with Colombian Cocaine Data

Eunseong Jang, University of Maryland, College Park, MD, United States, Margret Bjarnadottir, Marcus Boyd, Galia Benítez

Despite extensive efforts to curb cocaine trafficking, global cocaine production has increased for over a decade. Cocaine seizures have risen in tandem with production, but this has not translated into a reduction in cocaine consumption over the same period. This underscores the need for more effective strategies to enhance cocaine seizure

operations. Efficiently disrupting cocaine supply chains would begin with accurate predictions of trafficking routes. In this study, we aim to identify previously unrecognized trafficking nodes—both sources and destinations—by analyzing Colombian cocaine data. By examining the relationship between these observations and trafficking nodes inferred from anecdotal data, we seek to uncover unknown trafficking routes by narrowing the set of probable sources and destinations. Since trafficking nodes can be represented as a binary variable, we employ logistic regression to identify key factors that influence the likelihood of a location serving as a source or destination. However, applying a single regression model across all Colombian municipalities may be overly simplistic, as the cocaine trade is not controlled by a single entity. Consequently, certain factors may be associated with trafficking routes only in specific regions rather than nationwide. To address this issue, we employ a localized spatial analysis method—local geographically weighted regression (GWR), which only includes neighboring data when fitting the regression model of a focal point. Our results show geographical dynamic influence of factors associated with the trafficking nodes while overcoming limitations in the conventional GWR framework, which provide better understanding of clandestine cocaine network in Colombia.

4 - Optimizing Treatment Strategies for Small Renal Masses Using a Markov Decision Process Framework

Wendy Qi, University of Virginia, Charlottesville, VA, United States

Small renal masses (SRMs) are increasingly diagnosed due to widespread imaging, yet no consensus exists on optimal treatment strategies that balance cancer control and long-term renal function preservation. Clinical decisions are often made under uncertainty due to limited long-term data and variability in disease progression. Decision-making is further complicated by patient heterogeneity, tumor characteristics, and the evolving nature of tumor growth and kidney function decline.

To address these challenges, we develop a finite-horizon Markov Decision Process (MDP) model to simulate patient health trajectories and evaluate treatment strategies. The model incorporates key clinical factors, including tumor progression, chronic kidney disease (CKD) stages, recurrence, metastasis, and competing mortality risks. It evaluates treatment actions - active surveillance, ablation, partial nephrectomy, radical nephrectomy, and secondary treatment - based on their long-term effects on renal function, cancer control, and cause-specific mortality.

Using the PRISM model checker, we perform probabilistic model checking to compute optimal treatment policies based on expected cumulative rewards, reflecting quality-adjusted life years (QALYs) tied to each health state and action. The model supports patient-specific simulations from ages 65 to 85, allowing us to identify the timing and type of intervention that maximize clinical outcomes.

Our findings show that optimal strategies vary by patient characteristics, including age, sex, CKD stage, and tumor size. The proposed modeling framework supports personalized planning by allowing the objective function to be adapted to patient preferences or clinician priorities. This enables tailored recommendations that enhance shared decision-making between patients and providers and support improved long-term outcomes in SRM management.

5 - Predicting Substance use Risk in Youth: Associations Between Mental Health and Environment in ABCD Data

Viviana Milla Angeles, University of Virginia, crozet, VA, United States, Daniel Otero Leon

Early substance use is associated with long-term negative health and behavioral outcomes. This project examines how mental health conditions and environmental influences relate to early substance use among adolescents, where alcohol, nicotine, and cannabis are among the most used substances in this age group.

We use the comprehensive Adolescent Brain Cognitive Development (ABCD) dataset, which offers a robust, longitudinal dataset tracking biological, psychological, and social development in youth across the United States.

We propose a two-stage stacking prediction model that incorporates five years of longitudinal data that includes demographic variables, social context (including family dynamics, peer interactions, and socioeconomic indicators), and diagnosed mental health disorders to uncover patterns and estimate the likelihood of substance use initiation. For the base level, we employ a non-linear model which allows us to capture complex, non-linear relationships in the data. At the second level, a logistic regression meta-model, which assumes linearity and serves as a robust aggregator. Using a non-linear base model helps uncover intricate patterns, while the logistic regression meta-model reduces the risk of overfitting, especially when trained on out-of-fold predictions from the base models. This architecture enables the final output, probability estimates, to combine the strengths of diverse predictive models.

The goal of our project is to support early identification by recognizing which children are most at risk of initiating substance use. Our findings aim to inform public health strategies and clinical decision-making focused on youth mental health and substance use prevention.

6 - Polysubstance Overdose Deaths Involving Fentanyl

Hyojung Kang, University of Illinois at Urbana Champaign, Champaign, IL, United States, Moe Toyoda, Aaliyah Ali, Stanley Ratemo, Bryan Miller

Fentanyl continues to drive a significant portion of overdose mortality in the United States, often in combination with other substances. Employing toxicology records from medical examiner data in Illinois from 2016 to 2024, this study aims to identify frequently co-used substances with fentanyl and examine how these combinations vary over time, across demographic groups and by geographic regions. Using association rule analysis, we systematically explore the prevalence, correlations, and conditional probabilities of frequently co-used substances. The findings provide a comprehensive understanding of polysubstance use behaviors surrounding fentanyl-related overdoses and demonstrate the value of data mining techniques in public health surveillance. Also, results can inform targeted harm reduction strategies and guide resource allocation to address the evolving dynamics of the overdose crisis.

SA24

Building A Level 4 A410

Applications of Statistical Reinforcement Learning

Invited Session

Applied Probability Society

Chair: Kelly Zhang, Imperial College London, London, N/A

1 - Statistical Inference for Misspecified Contextual Bandits

Yongyi Guo, University of Wisconsin Madison, Madison, WI, United States

We study statistical inference following adaptive experiments in contextual bandit settings—a framework widely used in practice. Real-world environments often involve model misspecification, which can undermine both decision quality and the validity of post-hoc inference.

We develop valid inference procedures for data collected by adaptive algorithms, without assuming a correctly specified reward model. We characterize general conditions under which inference remains valid and highlight *policy convergence* as a key requirement. This condition is broad and holds for many online algorithms, but is notably violated by standard contextual bandit methods such as LinUCB and Thompson Sampling when they rely on a misspecified model.

Our results offer new insights into the interplay between model misspecification and statistical inference in adaptive decision-making, addressing common challenges posed by structural mismatch and limited data.

2 - Reinforcement Learning for Real-World Inventory Management

Omer Gottesman, Amazon, New York, NY, United States

Managing inventory across a large, dynamic supply chain presents a wide array of interdependent challenges. At Amazon, we develop reinforcement learning methods to help make better decisions for problems such as how much inventory to buy and where to place it across our network of fulfillment centers. Our systems must navigate complex trade-offs under uncertainty, including variable vendor lead times, capacity constraints, correlated customer demand, and the risk of lost sales. In this talk, we will highlight how reinforcement learning can enable us to address these challenges in a unified framework. We will share practical lessons from developing realistic simulators from historical data and discuss how we evaluate and deploy policies in high-stakes, real-world environments.

3 - Dynamic Contextual Pricing with Doubly Non-Parametric Random Utility Models

Jiayu Li, New York University, New York, NY, United States

In the evolving landscape of digital commerce, adaptive dynamic pricing strategies are essential for gaining a competitive edge. This paper introduces novel doubly nonparametric random utility models that eschew traditional parametric assumptions used in estimating consumer demand's mean utility function and noise distribution. Existing nonparametric methods like multi-scale Distributional Nearest Neighbors (DNN and TDNN), initially designed for offline regression, face challenges in dynamic online pricing due to design limitations, such as the indirect observability of utility-related variables and the absence of uniform convergence guarantees. We address these challenges with new population equations that facilitate nonparametric estimation within decision-making frameworks and establish the first analytical results on the uniform convergence rates of DNN and TDNN, enhancing their applicability in dynamic environments.

Our theoretical analysis confirms that the statistical learning rates for the mean utility function and noise distribution are minimax optimal. We also derive a regret bound that illustrates the critical interaction between model dimensionality and noise distribution smoothness, deepening our understanding of dynamic pricing under varied market conditions. These contributions offer substantial theoretical insights and practical tools for implementing effective, data-driven pricing strategies, advancing the theoretical framework of pricing models and providing robust methodologies for navigating the complexities of modern markets.

4 - Contextual Bandits for Enhancing User Engagement in the Drink Less App

Xueqing Liu, Harvard University, Boston, MA, United States

Mobile health (mHealth) interventions often seek to improve long-term outcomes by delivering timely, personalized support that optimizes short-term engagement. Contextual bandits provide a promising framework for such just-in-time adaptive interventions, yet their use in mHealth has been limited by challenges in modeling count-based outcomes. This study introduces methods that integrate count data models (Poisson, negative binomial, zero-inflated Poisson, and zero-inflated negative binomial regressions) with Thompson sampling to address this gap. Using data from the Drink Less trial, we demonstrate that these approaches can improve user engagement with the app. Results from simulated datasets further confirm performance gains, showing higher cumulative proximal outcomes compared to existing algorithms. We also provide theoretical results on regret bounds to support the robustness of the proposed methods.

5 - Jitai-Twins: a Digital Twin Framework for Reinforcement Learning Algorithms and Just-in-TIME Adaptive Interventions

Susobhan Ghosh, Harvard University, Cambridge, MA, United States, Asim Gazi

Recent work on just-in-time adaptive interventions (JITAI) -- nascent precision medicine systems that deliver just-in-time healthcare support during daily life that adapts to the changing state of an individual -- have leveraged reinforcement learning (RL) algorithms to optimize support. A challenge in designing JITAI that leverage RL algorithms is that RL algorithm design can require numerous non-trivial decisions that must be made before deploying a JITAI in the real world (e.g., hyperparameter selection). Additionally, each deployment can be costly, precluding the use of A/B testing for every design decision. How should design decisions be made strategically before (and between successive) JITAI deployments? In this talk, we will discuss our proposed framework, digital twins for just-in-time adaptive interventions (JITAI-Twins), to address this question. JITAI-Twins are "digital twins of a subpopulation" that are used to virtually simulate the potential outcomes of a JITAI's design decisions for an upcoming deployment. Based on simulation results, design decisions are made for the deployed JITAI. To continually improve the JITAI, data collected during deployment is used to update the JITAI-Twin -- and this bidirectional feedback between deployments and simulation environments continues. JITAI-Twins are thus fit-for-purpose instantiations of the digital twin concept. Importantly, JITAI-Twins highlight continuity over the course of a JITAI's optimization and continual improvement, emphasizing the need for bidirectional feedback between versions of a simulation environment and a JITAI's real-world deployments.

SA25

Building A Level 4 A411

Reliable and Resilient Facility Logistics

Invited Session

TSL: Facility Logistics

Chair: Ritesh Ojha, Amazon, Redmond, WA, 98052, United States

1 - Deep Reinforcement Learning for Joint Location-Inventory Problems with Unknown Demand in Mobile Retail

Jiayin He, Tsinghua University, Beijing, China, People's Republic of, Wei Qi

The retail sector is rapidly transforming, with autonomous mobile stores (MFs) emerging as a novel format combining mobility's convenience with physical retail's engagement, particularly potent in addressing service equity issues like food deserts. However, effectively managing a fleet of MFs presents significant challenges, primarily the need to make high-frequency, dynamic joint location-inventory decisions under uncertain spatiotemporal demand. Traditional methods often struggle with the scale, dynamism, and rich features inherent in this problem. This research addresses the joint location-inventory problem for a central retailer managing MFs serving dynamically partitioned urban areas (Voronoi blocks based on grids). Demand is uncertain but characterized by grid-level features. We propose a data-driven Deep Reinforcement Learning (DRL) framework, specifically employing an Actor-Critic architecture enhanced with a spatial-temporal attention mechanism inspired by the Transformer. This approach learns an end-to-end policy directly from simulated or historical data, capturing complex dependencies without restrictive assumptions. The attention mechanism is crucial for effectively processing high-dimensional spatial and temporal inputs to inform better replenishment and relocation actions for each MF. Our approach provides valuable enhancements and partially achieve superior performance compared to traditional benchmarks (e.g., base-stock, myopic policies, static planning) and baseline DRL methods in extensive numerical experiments, including reductions in operational costs, improved service and fairness across different service blocks, and demonstrated scalability. Beyond practical implications for retailers seeking efficient and adaptive MF management strategies, this work contributes academically by showcasing advanced DRL applications in dynamic logistics and offers insights into balancing economic efficiency with social equity in modern retail operations.

2 - Optimizing Facility Layout and Vehicle Fleet Decisions for the Strategic Assembly Line Feeding Problem

Hannah Verplancke, Ghent University, Ghent, Belgium, Veronique Limère, El-Houssaine Aghezzaf, Emmanouil Thanos

The Strategic Assembly Line Feeding Problem (SALFP) aims at designing a robust internal logistics system for manufacturing companies to carry out their operations in an efficient manner over a long period of time. The SALFP consists of two core subproblems: the Supermarket Location Problem (SLP) and the Transport Vehicle Selection Problem (TVSP). In addition, SALFP also considers the influence of the tactical operating conditions by incorporating the tactical Assembly Line Feeding Problem (ALFP) decisions dealing with part feeding policy assignment in the strategic framework. The SLP decisions include locating and sizing supermarkets where parts are prepared according to their assigned policy. Positioning these supermarkets closer to the warehouse or the assembly line will influence the transportation decisions made by the TVSP component. The TVSP decides on the number and type of transportation vehicles to be purchased by the manufacturing company, limiting the purchasing cost. Those vehicles will subsequently be utilized to transfer all parts from the warehouse to the assembly line while maintaining a feasible routing schedule. By integrating these two correlated problems, we aim at providing an overall optimized solution for the internal logistics in high variability and low volumes manufacturing companies. We evaluate the performance of the solutions based on multiple scenarios in order to create robust solutions that can be put to use for multiple years in practice.

3 - Robust Outbound Load Planning with Volume Splitting for Parcel Carriers

Ritesh Ojha, Amazon, Redmond, WA, United States

In this work, we focus on the middle-mile consolidation networks of large parcel carriers where a flow plan specifies the primary and alternate flow paths for the commodities in the network, and a load plan specifies the number of trailers to operate between hubs, to move packages between their origin and destination hubs through intermediate sorting hubs. The joint optimization of flow and load planning decisions in large-scale networks is intractable. Therefore, we consider a simpler subproblem focused on specifying a low-cost load plan outbound from a single hub, for a series of consecutive sorting periods, given pre-specified flow plans for commodities. As this problem is solved a few days before the day of operations, we propose the robust outbound load planning problem and model it as a two-stage robust optimization model with relatively complete recourse. We solve the robust problem using a master-subproblem framework embedded in an exact column-and-constraint (CCG) algorithm. We propose heuristic scenario generation ideas that leverage the structure of the problem to provide tight lower bounds to the optimal objective without having to solve the computationally intensive MILP formulation of the subproblem. We integrate these ideas with a mountain climbing heuristic to efficiently generate local optimal solutions to the computationally-intensive subproblem in the CCG algorithm. On average, the local optimal solutions are within 11%-17% of the best bound obtained at the root node of the branch-and-bound tree when solving the subproblem using a commercial solver. An extensive computational study on real-life data shows that having

4 - A Dynamic Location and Routing Problem for Same-Day Delivery Using Autonomous Modular Vehicle as Mobile Parcel Stations

Amir Shafiee, University of Illinois Chicago, Chicago, IL, United States, Hanieh Rastegar Moghaddam Bajestani, Mostafa Ameli, Jane Lin

The rapid expansion of e-commerce and customer expectations for faster and flexible deliveries have intensified the challenges in last-mile logistics. Traditional systems often result in high costs, traffic congestion, and inefficiencies under fluctuating demand. In this study, we introduce a new last-mile delivery system using Mobile Automated Parcel Stations (MAPS), operated by Autonomous Modular Vehicles (AMVs) or pods. These delivery pods can station and move themselves in the neighborhoods; they can also join and disjoin in real time, forming energy-efficient platoons that reduce travel costs during shared route segments. We develop a location and routing optimization model formulated as a mixed-integer program (MIP) to determine optimal MAPS placement and pod routing. The model considers both pre-scheduled and real-time delivery or pickup requests, incorporating time windows, customer walking distance thresholds, pod capacity, and platoon-based cost reductions. Initial routes are planned at the start of the day, and dynamically updated as new requests arrive. The system decides whether to reroute existing pods or deploy a new pod without requiring returns to the depot. The framework integrates predictive planning for known requests with flexible route adjustments to accommodate dynamically arriving same-day pickups and deliveries. In the numerical experiments, we analyze the benefits of modularity and mobile stationing in different locations. The results are compared against an idealized scenario in which all customer requests are available at the beginning of the day. This comparison highlights the value of AMV-enabled modularity and dynamic planning in narrowing the performance gap between real-world uncertainty and full-information scenarios.

SA26

Building A Level 4 A412

Location Analysis for Electric Vehicle Charging Under Uncertainty

Invited Session

Location Analysis

Chair: Xian Yu, The Ohio State University, COLUMBUS, OH, United States

Co-Chair: Huangrong Sun, The Ohio State University, Columbus, OH, United States

Co-Chair: Huangrong Sun, The Ohio State University, Columbus, OH, United States

1 - Maximum Flow-Based Formulation for the Optimal Location of Electric Vehicle Charging Stations

Miguel Anjos, University of Edinburgh, Edinburgh, United Kingdom

We consider the problem of optimal expansion of EV charging infrastructure to provide a greater quality of service in terms of user satisfaction (via the availability of charging stations). With a focus on urban areas, we first propose a model for the assignment of charging demand to stations, framing it as a maximum flow problem. This model is the basis for the evaluation of user satisfaction with a given charging infrastructure. Secondly, we incorporate the maximum flow model into a mixed-integer linear program, where decisions on the opening of new stations and on the expansion of their capacity through additional outlets is accounted for. We showcase our methodology for the city of Montreal, demonstrating the scalability of our approach to handle real-world scenarios. We conclude that considering both spatial and temporal variations in charging demand is meaningful when solving realistic instances. This is joint work with Pierre-Luc Parent, Margarida Carvalho, and Ribal Atallah.

2 - Contextual Distributionally Robust Optimization Under Multi-Modal Decision-Dependent Uncertainty with An Application to EV Charging Station Location

Huangrong Sun, The Ohio State University, Columbus, OH, United States, Beste Basciftci, Xian Yu

We study a two-stage electric vehicle (EV) charging station location problem under uncertain customer demand, where the demand may depend on the EV charging station location decisions and exhibit distinct patterns in different regions. To calibrate this decision-dependent and multi-modal uncertainty based on some historical data, we first formulate a mathematical optimization program to jointly optimize the number of modes, the assignment of data points to modes, and the regression function in each mode. Given the complexity of this optimization model, we apply cluster-wise regression algorithms to solve it more efficiently, which can converge to locally optimal solutions under mild conditions. Leveraging the estimated mode probabilities, regression functions, and the empirical residuals obtained from the regression step, we formulate a two-stage distributionally robust optimization (DRO) model for the EV charging station location problem with both moment-based and Wasserstein ambiguity sets and derive their tractable reformulations. We compare the performance of our cluster-wise regression algorithms with various regression benchmarks through extensive numerical experiments using synthetic datasets and real-world data from New York State. We also illustrate the benefit of considering decision-dependency and multi-modality by comparing our DRO model with decision-independent or single-modal counterparts.

3 - Large-Scale Location-Allocation Problem with Sinkhorn-Augmented Benders Decomposition

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

Electric vehicle (EV) adoption continues to accelerate despite significant uncertainties. In 2025, over 25% of global car sales is electric. To support this transition, the expansion of public charging infrastructure is essential, and determining optimal locations for charging stations is a critical and challenging task. This problem is typically modeled as a location-allocation problem. However, as the number of locations and demand scale, conventional solution methods become computationally infeasible, even when decomposition techniques are applied.

This work proposes an efficient approximation approach based on the Sinkhorn algorithm to address the allocation subproblem under dense-demand cases. The overall framework models the location-allocation problem as a two-stage mixed-integer nonlinear program, which is solvable through a tailored Benders decomposition scheme.

We provide a theoretical analysis of the proposed method, including (i) the validity of custom Benders cuts adapted to the entropic regularization introduced by the Sinkhorn approximation, and (ii) the convergence properties of the resulting decomposition algorithm.

Computational experiments demonstrate that our algorithm effectively solves large-scale instances where commercial solvers fail, highlighting its potential to support the strategic deployment of EV charging infrastructure and, ultimately, to facilitate broader EV adoption. This framework can also be used in other facility location problems with high demand, such as bike-sharing systems, micro-transit services, etc.

4 - Urban Electrification Policy in NYC

Daniel Vignon, New York University, Brooklyn, NY, United States, Yanchao Li, Prakrith Raja, Zhengtian Xu

While most EV charging occurs at home, this relies on private, dedicated parking—a luxury in dense, urban settings like NYC, where private spaces are scarce. The resulting inconvenience for current and prospective EV owners has spurred cities nationwide to install, mandate, facilitate, or subsidize charging stations in garages, residential units, and curbside spots. Yet these well-meaning policies carry effects beyond boosting EV uptake: they reshape when and where people travel, alter parking

search costs, and influence how parking operators set prices and allocate spaces. Assessing their full welfare impact therefore requires a model bringing together travel-mode choice, individual activity and charging behavior, and parking market dynamics.

We fill this gap with an empirically calibrated framework for NYC. First, we integrate mode choice—EVs versus gasoline vehicles or non-driving options—with activities and charging decisions. Second, we model parking operators' decisions on pricing and supply as a game constrained by drivers' activity and parking behavior. We address computational complexity by leveraging the problem's convex structure to compute the Nash equilibrium among operators efficiently. We also combine proprietary industry data on parking availability and prices with inversion techniques to recover key parameter distributions.

Our approach enables simulation of counterfactual electrification policies—ranging from curbside charging mandates to garage subsidy schemes—and predicts their impacts on parking search times, EV adoption rates, and space availability. These insights will help policymakers better allocated limited funds for EV adoption, ensuring that expansions both accelerate urban electrification and maintain efficient parking markets.

SA27

Building B Level 2 B201

Recent Advances in Integer and Low-Rank Optimization

Invited Session

Computing Society

Chair: Ryan Cory-Wright, Imperial, London, United Kingdom

1 - Improved Approximation Algorithms for Low-Rank Problems Using Semidefinite Optimization

Ryan Cory-Wright, Imperial, London, United Kingdom, Jean Pauphilet

Inspired by the impact of the Goemans-Williamson algorithm on combinatorial optimization, we construct an analogous relax-then-sample strategy for low-rank optimization problems. First, for orthogonally constrained quadratic optimization problems, we derive a semidefinite relaxation and a randomized rounding scheme, which obtains provably near-optimal solutions, mimicking the blueprint from Goemans and Williamson for the Max-Cut problem. We then extend our approach to generic low-rank optimization problems by developing new semidefinite relaxations that are both tighter and more broadly applicable than those in prior works. Although our original proposal introduces large semidefinite matrices as decision variables, we show that most of the blocks in these matrices can be safely omitted without altering the optimal value, hence improving the scalability of our approach. Using several examples (including matrix completion, basis pursuit, and reduced-rank regression), we show how to reduce the size of our relaxation even further. Finally, we numerically illustrate the effectiveness and scalability of our relaxation and our sampling scheme on orthogonally constrained quadratic optimization and matrix completion problems.

2 - The Augmented Factorization Bound for Maximum-Entropy Sampling

Yongchun Li, University of Tennessee Knoxville, Knoxville, TN, United States

The maximum-entropy sampling problem (MESP) aims to select the most informative principal submatrix of a pre-specified size from a given covariance matrix. This paper proposes an augmented factorization bound for MESP based on concave relaxation. By leveraging majorization and Schur-concavity theory, we demonstrate that this new bound dominates the classic factorization bound of Nikolov (2015) and a recent upper bound proposed by Li et al. (2024). Furthermore, we provide the theoretical guarantees that quantify how much our proposed bound improves the two existing ones and establish sufficient conditions for when the improvement is strictly attained. These results allow us to refine the celebrated approximation bounds for the two approximation algorithms of MESP. Besides, motivated by the strength of this new bound, we develop a variable fixing logic for MESP from a primal perspective. Finally, our numerical experiments demonstrate that our proposed bound achieves smaller integrality gaps and fixes more variables than the tightest bounds in the MESP literature on most benchmark instances, with the improvement being particularly significant when the condition number of the covariance matrix is small.

3 - The Surprising Performance of Random Partial Benders Decomposition

Yupeng Wu, London Business School, London, United Kingdom, Jean Pauphilet

Benders decomposition is a technique to solve large-scale mixed-integer optimization problems by decomposing them into a pure-integer master problem and a continuous separation subproblem. To accelerate convergence, we propose Random Partial Benders Decomposition (RPBD), a decomposition method that randomly retains a subset of the continuous variables within the master problem. Unlike existing problem-specific approaches, RPBD is simple to implement and universally applicable. We present both computational and theoretical evidence to support its effectiveness. For example, in extensive numerical experiments on network design and facility location problems, we find that (i) RPBD accelerates the convergence of Benders decomposition for problems with and without relatively complete recourse; (ii) RPBD usually halves the optimality gap at termination compared with a standard Benders approach; (iii) a random retention strategies is just as effective as problem-specific approaches proposed in the literature.

4 - Implicit Regularization of Infinitesimally-Perturbed Gradient Descent Toward Low-Dimensional Solutions

Jianhao Ma, University of Pennsylvania, Philadelphia, PA, United States, Geyu Liang, Salar Fattahi

Implicit regularization refers to the phenomenon where local search algorithms converge to low-dimensional solutions, even when such structures are neither explicitly specified nor encoded in the optimization problem. While widely observed, this phenomenon remains theoretically underexplored, particularly in modern over-parameterized problems. In this paper, we study the conditions that enable implicit regularization by investigating when gradient-based methods converge to second-order stationary points (SOSPs) within an implicit low-dimensional region of a smooth, possibly nonconvex function. We show that successful implicit regularization hinges on two key conditions: (i) the ability to efficiently escape strict saddle points, while (ii) maintaining proximity to the implicit region. Existing analyses enabling the convergence of gradient descent (GD) to SOSPs often rely on injecting large perturbations to escape strict saddle points. However, this comes at the cost of deviating from the implicit region. The central premise of this paper is that it is possible to achieve the best of both worlds: efficiently escaping strict saddle points using infinitesimal perturbations, while controlling deviation from the implicit region via a small deviation rate. We show that infinitesimally perturbed gradient descent (IPGD), which can be interpreted as GD with inherent "round-off errors", can provably satisfy both conditions. We apply our framework to the problem of over-parameterized matrix sensing, where we establish formal guarantees for the implicit regularization behavior of IPGD. We further demonstrate through extensive experiments that these insights extend to a broader class of learning problems.

SA28

Building B Level 2 B202

ORSTW

Invited Session

INFORMS Taiwan Chapter

Chair: I-Lin Wang, National Cheng Kung University, No.1 University Rd., Tainan, 701, Taiwan

1 - Expectation, Risk-Averse, and Distributionally-Robust Optimization for the Real-World Challenges in Emergency Medical Services

Yu-Ching Lee, National Tsing Hua University, Hsinchu, Taiwan

Emergency Medical Services (EMS) systems face significant operational challenges due to inherent uncertainties in demand, resource availability, and response times. Traditional optimization models often fall short in addressing these complexities, particularly when considering the variability and risks associated with real-world EMS operations. This research studied the frameworks of three key approaches: expectation optimization, risk-averse optimization, and distributional-robust optimization, to improve decision-making in EMS under uncertainty. The expectation optimization approach, or the stochastic programming, aims to minimize the average response time or cost while managing resource allocation across multiple service scenarios. The risk-averse optimization framework introduces tools to minimize the worst-case

outcomes, catering to situations where high-risk scenarios (e.g., critical patient cases) require robust and fail-safe planning. Lastly, distributional-robust optimization addresses uncertainty in both the probability distribution of future demands and service parameters, without assuming a specific underlying distribution, providing more flexibility and reliability in complex environments.

2 - Analyzing Manufacturer Channel Encroachment and Retailer Service Strategies Under A Carbon Emissions Trading Scheme

Jen-Yen Lin, National Cheng Kung University, Taiwan, East Dist., Taiwan

With the global push toward carbon neutrality, enterprises are facing increasingly stringent carbon reduction regulations. Achieving environmental sustainability while maintaining operational efficiency has become a critical issue in supply chain management. This study investigates a two-tier sustainable supply chain composed of a manufacturer and a retailer and develops a decision-making model that incorporates a carbon emissions trading scheme. In the proposed model, the manufacturer can sell directly to end consumers through an online platform, while the retailer responds to the resulting competition by enhancing its service level. The study further analyzes the optimal pricing and operational strategies for both the manufacturer and the retailer under this setting.

Three decision-making scenarios are analyzed: (1) the manufacturer does not engage in direct sales and the retailer does not enhance service (ND); (2) the manufacturer does not engage in direct sales while the retailer improves service (NS); and (3) the manufacturer engages in direct sales while the retailer does not improve service (ED). A Stackelberg game framework is employed to derive the optimal pricing strategies and corresponding profit for both players under each scenario. Numerical results reveal that when consumers exhibit a high psychological discount toward direct channels, the manufacturer tends to adopt a direct-selling strategy, and the advantage of doing so increases as the unit carbon emission decreases. Conversely, when the carbon trading price is relatively low or consumers perceive only a minor discount toward direct channels, refraining from direct sales may yield higher profits for the manufacturer.

3 - XGBoost with Missing Data Penalty for Multimodal Alzheimer's Disease Classification

Chen-ju Lin, Yuan Ze University, Taoyuan, Taiwan, Yu-chen Chiu

Multimodal clinical datasets are common in Alzheimer's disease (AD) diagnosis but often contain missing data due to incomplete testing or patient dropout. Traditional strategies, such as listwise deletion or imputation, may degrade classification performance, especially when the missingness mechanism is complex. This study proposes a new objective function for the XGBoost algorithm by incorporating the penalty of having missing data. Feature weights are iteratively optimized to minimize a composite loss function that reflects prediction performance, regularization, and data presence. The method was evaluated on a cohort of 517 subjects, comprising four data modalities: demographics together with APOE4 gene and cerebrospinal fluid biomarkers, magnetic resonance imaging (MRI), cognitive test scores, and the neuropsychological assessment battery. Compared to standard XGBoost, the proposed approach achieved higher F1 scores, improving by 0.02, 0.09, and 0.11 as the missing data rate increased from 20% to 30% and 40%, respectively. Feature importance analysis indicated MRI as the key predictor. This work demonstrates that augmenting XGBoost with a tailored penalty structure that models missingness in the learning objective enhances classification performance. The proposed algorithm offers a robust solution for healthcare applications with incomplete multimodal data.

4 - Minimizing Worst-Case Forest Fire Damage Through Optimal Prescribed Burning Strategy

I-Lin Wang, National Cheng Kung University, Tainan, Taiwan

Forest fires threaten ecosystems, with over 51,320 incidents burning 8 million acres in 2024. Prescribed burning mitigates wildfire risk, but optimizing burn locations remains a challenge. We model this as a Stackelberg game using bilevel programming, where the forest manager selects burn sites while an adversary chooses ignition points to maximize damage. Our approach considers worst-case scenarios, incorporating fire spread dynamics and cost-effective treatment areas. By reformulating the model into a single-level integer program, we improve efficiency and develop heuristics for faster solutions. Numerical results show our strategy significantly reduces fire damage, enhancing wildfire defense and broader risk mitigation planning.

SA29

Building B Level 2 B203

Optimizing Distributed Energy Generation Systems

Invited Session

ENRE: Energy

Chair: Alexandra Newman, Colorado School of Mines, Golden, CO, United States

1 - Sizing of Stand-Alone Residential Battery Energy Storage Systems

Andrew Aikman, Colorado School of Mines, Golden, CO, United States, Karlyle Munz, Paulo Tabares, Alexandra Newman

Lithium-ion battery energy storage systems (BESS) can provide cost and CO₂ emissions savings to residential stakeholders by shifting energy usage from peak to off-peak hours. Lifetime cost savings can now be achieved without co-located generation due to the emergence of storage-based incentive programs. This work offers two methodologies for the sizing of stand-alone residential BESSs. Our life-cycle-cost approach uses incentive programs to balance upfront capital expenditures against lifetime electricity cost savings. The iterative-heuristic approach adds modules until the marginal benefit reaches a specified tolerance. We consider five control strategies to address both electricity cost and emissions goals. In a case study, we select systems for 29 low-income homes located in Lake County, Colorado, based on electrified and retrofitted electric demand profiles and local grid emissions rates. We isolate the cost and CO₂ emissions effects of electricity pricing by applying structures from across the United States. Both sizing methodologies achieve electricity cost savings of -50 to 40% and emissions savings of -12 to 23%. In another study, we create electric load profiles for the Lake County homes as if they were relocated to California, New York, Texas, Florida, and Nebraska. The new electric loads are paired with local emissions rates and pricing structures to create real home-pricing-emissions combinations. These cases produce electricity cost savings of -50 to 45% and emissions savings of -20 to 72%. Cost savings are highest in locations with pricing structures for electrified buildings while emissions saving potential depends on annual energy usage and grid-based emission rates.

2 - Minimizing Unmet Demand in Small-Volume Heat Pump Water Heaters

Janelle Domantay, School of Mines, Golden, CO, United States

Heat pump water heaters leverage the heat from ambient air temperature to heat water without using electric resistors. This method is up to four times more efficient than the commonly used electric water heaters and have been proposed to reduce energy costs. It is necessary to manually indicate a heating setpoint that determines the minimum temperature to maintain in the internal tank. However, depending on water usage and setpoint temperature, the hot water outlet temperature can fluctuate greatly. In order to maximize the cost efficiency of heat pump water heaters (HPWH), it is desirable to minimize the setpoint while still maintaining comfortable hot water outlet temperatures. Here we show that based on 220 days of minute-level draw data from a single HPWH site, we can accurately predict outlet temperatures for additional sites within a 2°C (1.8°F) margin of error across a two hour prediction horizon. This level of forecasting allows us to maximize the amount of delivered hot water for 120 V HPWHs and capitalize on variable electricity prices by maximizing preheating during times of the day with the lowest electricity demand.

3 - Optimizing Design and Dispatch of a Microgrid with Heterogeneous Heating Loads and Air-Sourced Heat Pumps

Alexander Zolan, National Renewable Energy Laboratory, Golden, CO, United States

The adoption of distributed energy resources to serve all or part of a building's electrical load is growing in response to the declining cost of onsite renewables and energy storage. In addition to this, electrification of the built environment motivates the inclusion of thermal loads when considering technology selection to meet a building's combined energy needs. We introduce REopt, an open-source web tool that obtains cost-minimizing distributed energy resource designs subject to meeting electrical, heating, and cooling loads with or without a utility connection for residential and commercial customers. We present an extension to include multiple heating loads of

different quality and the availability of air-sourced heat pumps as an option for meeting either space heating or domestic hot water needs. We demonstrate the impact of these extensions on design selection. We then present case studies showing the impact of technology availability, costs, and emissions reduction goals on technology selection.

4 - Enhancing Energy Resilience on Military Installations: a Heuristic Optimization Approach Using Genetic Algorithms

James Grymes, United States Military Academy, West Point, NY, United States, Jake Moffat

Functioning energy infrastructure on military installations is essential in executing daily operations. From powering ranges to maintaining communication systems, energy is a driving factor, and a failure has dramatic consequences. The research conducted by the Energy Resilience team at the United States Military Academy is directed toward enhancing the resilience of the Army's energy infrastructure through a heuristic optimization approach. Deliverables include innovative, sustainable, and actionable design solutions that address resilience, cost, and emissions. Improving resilience, or how quickly a system can return to a satisfactory state after a failure, is the primary focus. The research presented will be the analysis of the heuristic optimization method, recommendations to the algorithm, and current progress. The model incorporates a genetic algorithm to produce a Pareto frontier of micro-grid configurations. The genetic algorithm is coded in MATLAB and works to optimize the defined fitness function. The current results of the model analysis have led to recommendations for implementing adaptive mutation and elitism and adding resilience to fitness functions. The model produced will provide insight into improving energy infrastructure on military installations with a focus on resilience.

5 - Design of a 200-Kw Scale Solar Receiver Test Facility Using An Available Tower

Sebastian Dogue, Colorado School of Mines, Golden, CO, United States, Andrew Aikman, Alexandria Newman, Gregory Jackson

The scale-up of concentrating solar technologies requires adequate on-sun test facilities at or above the 100-kWth scale to demonstrate components and economically evaluate new concepts at plant-scale performance conditions. Adaptation of available tower structures provides a low-cost pathway to satisfy this need. This study examines the possible conversion of an existing 60-m high tower facility in the Front Range region of Colorado to a 200-kWth scale solar receiver and heliostat field test facility. We use NREL's SolarPILOT to examine the viability of the tower, surrounding environs, and weather to offer up to 500 kW of thermal power over enough days per year to support effective research and development. We use two sizing approaches to generate heliostat fields which achieve the desired mean receiver flux and uniformity. The 200-kW layout method creates a densely packed field of heliostats with focused aiming strategy, and the max-power layout method selects more heliostats but uses a dispersed aiming strategy. The max-power layout method obtains much greater flux uniformity at the expense of optical efficiency, which is necessary for receiver demonstration tests. Available receiver locations on the top half of the tower offer a suitable range of conditions to test 1x1m receivers with mean fluxes of 250 kW/m², from which final location can be selected based on demonstration test objectives.

SA30

Building B Level 2 B204

Navigating Future Uncertainties in Energy System Planning

Invited Session

ENRE: Electricity

Chair: Sheng Lun Cao, University of Calgary, Calgary, AB, Canada

1 - Representative Period Selection for Incorporating Future Weather Data in Capacity Expansion Modeling

Sheng Lun Cao, Carnegie Mellon University, Pittsburgh, PA, United States

Long-term capacity expansion planning needs to account for future weather uncertainties to ensure renewable resource adequacy and grid reliability. Global climate modeling can forecast future weather dynamics to the end-century, but capacity expansion models cannot leverage the full spatiotemporal data resolution due to computational tractability. Thus, representative period selection become a necessary step to select some future weather periods to simulate "typical" grid operations in the future. We evaluated four hard and fuzzy k-medoids clustering-based and one tree-based representative period selection methods to assess their impact on the final capacity expansion planning outcomes for the ERCOT grid in Texas during mid-century (2045 – 2054) and end-century (2085 – 2094). Through 1120 capacity expansion cases, we determined that the cluster size-weighted hard k-medoids clustering method (UHKM) is the best-performing representative period selection method, in terms of estimated total system cost. The fuzzy clustering methods systematically overbuilds wind capacity and underbuilds PV capacity by selecting representative periods with high wind and low PV availability. Representative period selection, despite being one step in capacity expansion modeling, can have a significant financial impact on long-term system planning, leading up to \$7.2 billion in total system cost underestimation.

2 - Explaining Variation in Emissions and Health Impacts from SO₂ Control Technologies Installed at U.S. Coal Power Plants

Joanna Slusarewicz, Carnegie Mellon University, Pittsburgh, PA, United States, Valerie Karplus, Nicholas Muller

When predicting outcomes of future environmental policy in the electricity sector, analysts must make assumptions about investment in and use of adopted technologies that may critically determine ex-ante estimates of costs to households and regulated firms as well as the resulting benefits to society. To illustrate the importance of such assumptions to environmental policy analysis, this work examines U.S. coal power plant operational decisions that affect the SO₂ removal efficiency of flue gas desulfurization (FGD) units in response to environmental and market regulation. It has often been assumed that plants installing FGDs will reduce pre-FGD SO₂ emissions by a factor of the equipment's rated removal rate. However, plants often have financial incentives to (a) burn cheaper, higher sulfur coal after installing SO₂ scrubbers, and (b) operate outside of the standard conditions under which pollution control equipment is rated. This study compares the effects of each behavior by decomposing SO₂ emissions as a function of coal sulfur content and SO₂ removal efficiency across plants from 2008 to 2021. The study finds that SO₂ emissions declined by 90.0% after FGD installation relative to the 95.6% predicted from the rated removal rates. This discrepancy is in part due to a 58% increase on average in percent sulfur by weight of coal combusted post-FGD installation. Furthermore, total SO₂ tonnage capture rates were 2.4% lower than predicted from the rated removal rates, resulting in 740 additional deaths per year on average as estimated by the Air Pollution Emissions Experiments and Policy version 4 (AP4) model.

3 - A Comparative Evaluation of Machine Learning Algorithms for Disaggregating Outages from Disconnections in Smart Meter Data

Akua McLeod, Carnegie Mellon University, Pittsburgh, PA, United States, Destenie Nock, Amritanshu Pandey

Electricity service disruptions compromise households' ability to meet energy demand and thus increase energy insecurity. Disruptions can take the form of power outages, or involuntary disconnections due to nonpayment. To achieve equitable energy access, guide targeted grid modernization efforts, and facilitate utility performance regulation, we must understand how electricity reliability varies spatially and across demographics. Advanced metering infrastructure (AMI) data offers a unique opportunity to evaluate electricity access at the ZIP code level, in the absence of utility published datasets, however we must first disaggregate periods of zero energy consumption into outages and disconnections. Building on our previous work which presented a novel method of classifying disruptions in AMI data using logistic regression, we compare the performance of several machine learning methodologies, using the Commonwealth Edison territory in Illinois as a case study.

To identify the superior classifier, we evaluate three models (logistic regression, a support vector machine and a neural network) using both standard accuracy metrics such as the F1 score and Cohen's Kappa, as well as by validating our aggregate results against known disruption statistics for the service territory. Leveraging our model results, we find that the most outage-burdened regions are concentrated in rural counties, and that regions with a higher percentage of low-income residents and people of color often carry higher outage burdens than the rest of the service territory. Ultimately, this work enables more granular metrics for electricity reliability, highlighting gaps in energy accessibility and informing the identification of energy-insecure communities.

4 - Mortality and Distributional Effects of Air Pollution from Integrated Steelmaking in the United States

Judy Park, Carnegie Mellon University, Pittsburgh, PA, United States, Nicholas Muller, Valerie Karplus

32% of iron and steel production in the US comes from Integrated Iron and Steel (IIS) plants, which, along with the metallurgical coke plants upstream of them, are major sources of air pollution. We modelled the atmospheric PM_{2.5} concentrations attributable to these plants' pollution emissions and estimated the magnitude of their premature mortality effects, plus their distribution across racial/income groups. Annual Scope 1 emissions of five PM_{2.5}-precursor pollutants were determined using data from all existing IIS plants (n=9) and metallurgical coke plants (n=13) in the US in 2017, and Scope 2 emissions were modeled using the average of projected electricity generation emission factors from 2026-2065 (reflecting a 40-year plant life). Atmospheric dispersion and secondary pollutant formation were modeled with a Reduced-Complexity Model (AP4T) to determine the resulting ambient PM_{2.5} concentrations at a census-tract resolution, from which the premature mortality risk was calculated using age-differentiated dose-response functions from a large cohort study. Results show that air pollution from IIS and coke plants in the US may be responsible for 700 deaths per year, equivalent to damages of \$7 billion/year (\$2020) or 30% of the total value of steel produced with integrated steelmaking. 333 of these deaths occur in the top 10% of most affected tracts (by pollution level), where the population-weighted mortality risk is 11.6 per 100,000 people, and Black, Other Nonwhite, AAPI, and Hispanic residents face higher-than-average mortality risks. Additionally, mortality risks increase slightly with increasing household incomes, especially for residents who are Hispanic, Other Nonwhite, or 2+ Races.

5 - Modeling Spatial Health Damage Disparities from CCS Under Atmospheric and Policy Uncertainty

Nexus Attiogbe, UMass Amherst, Amherst, MA, United States, Erin Baker, Paola Furlanetto, Michael Ash

Recent U.S. policies, including the Inflation Reduction Act's 45Q Tax Credit and the EPA's CO₂ Performance Standards, aim to incentivize carbon capture (CC) in the power sector. This study evaluates how these individual policies and their combinations affect CC deployment, co-pollutant emissions (NO_x, SO_x, PM_{2.5}, NH₃, and VOCs), and the distribution of health damages across socioeconomic groups. Using a stylized three-node model, we optimize the generation investment and dispatch. We apply this regional case studies in Alabama and Texas to investigate health disparities. Air pollution exposure modeling is paired with census tract-level data to assess disparities in per-capita monetized health damages. Regional variation in co-pollutant dispersion patterns due to atmospheric conditions is incorporated to reflect climate-based uncertainties. This study highlights the importance of considering environmental injustice when designing policy by analyzing trade-offs between CC adoption, co-pollutant exposure, and equity outcomes. Results emphasize the need for decarbonization strategies that remain robust under uncertainty and prioritize equitable emission reductions.

SA31

Building B Level 2 B205

Corporate Sustainability and ESG

Contributed Session

Contributed

Chair: Andy Buschmann, University of Michigan, Ann Arbor, Ann Arbor, MI, United States

1 - Environmental Shareholder Proposals and Firm Strategic Responses

Marilyn Lucas, University of Vermont, Burlington, VT, United States, Hung Do, Thomas Noordewier

We develop an economic model to study how firms decide on R&D and marketing investments in response to environmental shareholder proposals. To complement our analytical findings, we conduct an empirical study, using a dataset that combines financial data and shareholder resolution information.

2 - Impact of Epr on Circular Strategy for Electric Vehicle Batteries: Recycle or Refurbish?

Vikash Kumar, Ivey Business School, London, ON, Canada, Jason Nguyen

We study how implementation of extended producer responsibility (EPR) legislation for green products, such as EV batteries may impact their economic and environmental outcomes. EV batteries are critical to viability of EVs, which in turn is critical to meeting the emission targets. However, these EV batteries are equally environmentally detrimental to make as well as dispose. EPR regulations for EV batteries are policy tools that force EV manufacturers to choose different circular economic (CE) strategies to implement EPR standards. However, conspicuously there are no studies in academia that explores the unique nature of EV batteries, which make designing EPR implementation challenging. We develop an analytical model to analyze the different CE strategies of EV manufacturers under EPR and analytically reveal its economic and environmental impact. The impacts of EPR are not intuitive for EV batteries, on one hand it reduces the environmental degradation through recycling by mandating minimum proportion for input substitution from virgin material to recycled material; however, on the other hand this reduces the reuse of EV batteries which in turn aggravates environmental degradation. Furthermore, it may also reduce new production of EVs, which in turn slows down transition to green transportation, impacting the environment negatively. We offer insights into how the EPR regulations impact the environmental and economic welfare and reveal parameters of interest that the policy makers may target to nudge the EV manufacturers towards the optimal outcome.

3 - Modeling the Interplay Between ESG Dynamics and Corporate Financial Decisions

Jingwei Chen, University of Tennessee, Knoxville, Knoxville, TN, United States, Mingzhou Jin

As Environmental, Social, and Governance (ESG) considerations become central to corporate finance, decision-makers face growing challenges in aligning long-term sustainability goals with financial performance. This research develops a multi-period optimization framework that captures the dynamic interaction between ESG trajectories, investment portfolios, and debt financing decisions. The model integrates ESG score depreciation, asset deterioration, project-level sustainability impacts, and ESG-linked financing conditions, enabling firms to navigate both market-driven expectations and regulatory ESG targets. Each investment option is characterized by its financial return, long-term asset contribution, and ESG performance, and must be evaluated under liquidity and leverage constraints. Interest rates are endogenously defined as a function of firm size, leverage, and ESG score, linking sustainability performance to the cost of capital. The model tracks firm-level ESG scores and financial health over time and supports scenario analysis on strategic investment behaviors. Early-stage results from simulation-based evaluation highlight trade-offs between ESG impact and financial efficiency, and reveal how ESG-oriented capital structures affect long-term asset growth and interest rates. By incorporating ESG dynamics into corporate financial decision-making, this work aims to support transparent, robust, and actionable planning tools for firms transitioning toward sustainable operations. The proposed model also offers a foundation for future research exploring ESG-related financial incentives, policy instruments, and portfolio classification.

4 - Integration of Circulatory and Innovation Investments for Consumer Electronics Companies

Ece Zeliha Demirci, TED University, Ankara, Turkey, Emre Nadar, Zeynep Ece Livaoglu

We consider electronics companies that aim to allocate their limited financial resources along two critical dimensions: investments for research and development (RD) to improve the new product design and customer base, and investments for circularity technologies to increase the sales of remanufactured items. The volume of used items collected for remanufacturing can be controlled by the acquisition price set for the used item. We formulate the problem of such an electronic company with a bilevel programming model that takes into account demand uncertainty and supply-loop constraints. We characterize the optimal solution structure and derive managerial insights on how such companies can strategically allocate their resources.

5 - Does Corporate Sociopolitical Activism Oriented Market Exits Benefit Corporate Brand Equity and Firm Profitability? Evidence from Corporate Exodus from Russia Amidst the 2022 Russo-Ukrainian War

Wenjun SHU, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Di Fan, Xiao Zhang, Maggie Dong

Problem definition: With the rise of political consumerism, stakeholders increasingly pressure firms to take stands on sociopolitical issues. In response, multinational corporations (MNCs) engage in corporate social activism (CSA), such as publicly supporting or opposing sociopolitical issues. During the 2022 Russo-Ukrainian War,

many MNCs withdrew from the Russian market as a form of CSA, differing from traditional economically driven market exits and raising questions about their impact on firm performance. **Methodology/results:** We investigate whether CSA-driven market exits benefit firm performance in terms of corporate brand equity and profitability. Focusing on U.S.-listed MNCs that exited Russia during the 2022 Russo-Ukrainian War, we conduct a quasi-natural experiment comparing firms that announced withdrawal with those that did not. Our analysis reveals significant increases in corporate brand equity and profitability following the CSA decisions. Furthermore, the benefits are amplified when the firms' actions align with target market values and when social media sentiment is more positive. We address potential endogeneity issues through additional robustness tests, confirming the stability of our findings. **Managerial implications:** Our findings offer valuable insights for MNCs operating in global markets on how CSA-oriented market exits may affect their brand and profitability. The study provides guidance on effectively managing market exits amidst complex sociopolitical issues. Additionally, our results can inform policymakers about the quantitative effects of CSA-driven market exits and their potential economic impact, serving as important references for future policy-making.

6 - Do Firms Walk the Circular Talk? Multimodal Signals, Supplier Trust, and Dynamic Collaboration in the Circular Economy

Nan Liu, University of Birmingham, Birmingham, United Kingdom, Yufeng Zhang, Wanhong Li

As the Circular Economy (CE) becomes increasingly integral to sustainable operations, firms face intensified pressure to credibly signal their green commitments and secure long-term supplier collaboration. Traditional ESG disclosures provide retrospective insights with limited predictive capability, necessitating more forward-looking approaches in supply chain management. Building on signaling theory and organizational trust theory, we examine how firms' **multimodal data consistency**—integrating verbal, vocal, and visual expressions, dynamically shapes supplier trust formation and subsequent collaboration behaviors within sustainable supply chain contexts. We propose a **human-AI hybrid** methodology combining multimodal analytics with dynamic **agent-based modeling (ABM)**. Corporate executives' CE-related speeches are analyzed using multimodal tools (e.g., Whisper, openSMILE, OpenFace), generating **Multimodal Congruence Scores (MCS)**. Supplier trust perceptions and collaboration intentions are assessed through scenario-based surveys, embedding CE-specific constructs. To validate MCS, we implement human-AI alignment testing via the *Epitome* platform, where real suppliers rate trustworthiness of identical video content. Robust alignment between human and AI-derived ratings confirms MCS validity. Subsequently, we implement an **ABS** featuring adaptive supplier–firm interactions, dynamically capturing reciprocal trust updates and strategic communication adjustments. We pioneer the integration of multimodal expression congruence into the supply chain management literature, advancing signaling theory by redefining corporate signals as dynamic and interactive rather than static cues. It expands organizational trust theory by illustrating trust formation as an adaptive, feedback-driven process between firms and suppliers. Practically, our framework provides enterprises with scalable tools to strategically align organizational communication practices, thereby fostering resilient, trust-based supplier partnerships essential for sustained competitive advantage in circular supply chains.

7 - When Does Hedging Backfire? How Corporate Political Connections Affect Firm Survival Following Democratic Breakdown, Insights from the 2021 Coup in Myanmar

Andy Buschmann, University of Michigan, Ann Arbor, Ann Arbor, MI, United States

Firms frequently engage in political hedging—cultivating ties with competing political actors—to mitigate risks arising from regime turnover or policy shifts. While such strategies have been found effective in consolidated and new democracies, this study argues that political hedging may backfire in transitional regimes where democratic institutions remain fragile and the autocratic elite is still politically powerful. In such contexts, finite resources and costly commitments make political hedging not only more contentious but also potentially dangerous. When democratic transitions stall or reverse, autocratic elites may interpret ties to opposition actors as disloyalty or betrayal, triggering exclusion from strategic resources or political repression. To investigate this argument, I analyze firm survival following the 2021 military coup in Myanmar, which abruptly reversed a decade-long democratic transition. Based on data from a large-scale data leak of Myanmar's corporate registry and various other original datasets on firms' political ties, I explore the quasi-causal effect of the coup on firm survival conditional on different pre-coup corporate political connections (CPC). Within a factorial DiD framework, I classify firms by their network distance to opposing political camps prior to the coup. I supplement my causal analysis with data on post-coup repression to distinguish firm death due to market vis-à-vis non-market mechanisms. This study contributes to the nonmarket strategy literature by identifying critical boundary conditions under which political hedging ceases to function as an optimal strategy. I highlight the distinct political hazards firms face when pursuing CPC in politically uncertain regimes.

SA32

Building B Level 2 B206

Advances in Power Systems, Energy Market, and Sustainability

Invited Session

ENRE: Environment and Sustainability

Chair: Zuguang Gao, University of California, Irvine, Irvine, CA, 92697, United States

1 - Unlocking Demand-Side Flexibility with Dynamic Operating Envelope Market and AI Agent

Cong Chen, Cornell University, Ithaca, NY, United States

The global energy transition is transforming power systems through the rapid growth of renewables, data centers, and electrification across transportation, infrastructure, and industry. This talk highlights two critical challenges and our proposed solutions. First, distributed energy resources (DERs)—such as rooftop solar, home batteries, and electric vehicles—along with large new loads like data centers, require reliable and fair access to the distribution network. We introduce a dynamic operating envelope market, implemented through a forward auction, to allocate network injection and withdrawal rights. This mechanism enhances grid reliability while enabling equitable participation of diverse resources. Second, unlocking demand-side flexibility requires energy service providers to better understand customer behaviors and preferences in order to design sustainable and profitable business models. We develop AI agents powered by large language models that act as *homo silicus* for energy customers. These agents generate credible and diverse behavioral insights and capture responses to rare but high-impact events such as blackouts. Together, these approaches address renewable uncertainty, maintain competitive pricing, and advance demand-side participation in modern electricity markets.

2 - Reliability-Incorporated Optimization in Power Distribution via Multi-Source Weather Data and Ensemble Models

Gejia Zhang, Rutgers University, Piscataway, NJ, United States, Robert Mieth

Maintaining the reliability of modern power distribution systems is crucial, yet it is increasingly challenged by rising electricity demand, aging infrastructure, and the growing impact of unstable weather conditions. Recent studies have shown that operational decisions and environmental factors both play roles in determining component failure rates. This study integrates framework for enhancing power distribution system reliability by incorporating an array of meteorological factors from multiple measurement sources. To manage the resulting collinearity and computational challenges, we propose two ensemble tree-based models that systematically identify the most influential features and reduce the dataset's dimensionality based on each feature's impact on failure rate prediction. By weighting features according to their importance, critical variables exert a stronger influence on predicted failure rate, while each tree in the ensemble incorporate distinct feature subsets. These predictions are embedded within a sequential, non-convex optimization procedure, where operational decisions are dynamically updated in response to system states. We perform a numerical experiment that demonstrates the model's ability to integrate accurate reliability estimates derived from multi-dimensional data across multiple sources while co-optimizing operational costs.

3 - Data Driven Evaluation of Distribution Grid Upgrade Costs Under Utility Scale Smart Charging Strategies

Lusha Wang, University of Alabama, Tuscaloosa, AL, United States

Transportation Electrification will significantly reduce carbon emissions and reliance on fossil fuel. The world has seen an accelerating transition to electric vehicles (EVs) driven by environmental concerns, government promotions and battery technology advancements. On the one hand, this transition brings in great positive environmental impacts; on the other hand, it introduces enormous challenges to existing power systems which were designed and constructed without considering massive EV integration. It is thus imperative to understand the impact of EV charging on real-world power distribution networks and rethink the planning and operations of power grids. This talk will share key insights from a real-world smart charging management pilot program, evaluating strategies for optimizing EV charging loads and enhancing grid resilience. Leveraging real charging data from over 4,600 EV drivers and more than 2000 real-world distribution feeders, the study provides detailed feeder-level

analysis, revealing how factors like transformer rating and number, load profiles and infrastructure constraints influence smart charging performance. Data-driven methods are utilized to scale feeder-level insights to utility-wide estimates, yielding rapid yet precise results. These findings offer valuable guidance for stakeholders - including governments, utilities, third-party operators and consumers - on the role of feeder-specific smart charging in infrastructure planning and grid modernization.

4 - Energy Consumption of Limited Response to Weather

Fariba Farajbakhsh Mamaghani, Tulane University, River Ridge, LA, United States

The integration of smart technologies into electricity grids has enabled utilities to monitor consumption in real time and implement demand response (DR) programs that rely on dynamic pricing to balance supply and demand. Yet, recent experiences in the residential sector show mixed effectiveness, calling into question the assumption that households reliably adjust consumption in response to price signals. A key challenge is that prices must be set in advance and cannot adapt to random environmental factors such as weather. As a result, households often demonstrate limited ability to rationally modify their usage, leading to volatile demand patterns.

We address this gap by developing a novel framework that models households' decision-making under cognitive constraints, drawing on the concept of "rational inattention." Our model characterizes optimal consumption strategies that explicitly account for the trade-off between information processing costs and potential savings. This framework captures how random shocks in ambient conditions map into household consumption and identifies conditions under which households are responsive or unresponsive to demand fluctuations. We show that consumption is shaped by competing comfort-seeking and cost-saving incentives, yielding nontrivial demand patterns. Importantly, while higher prices reduce average demand, they may also increase sensitivity and variability. Likewise, reducing cognitive costs may unintentionally raise variability if prices are not set appropriately.

Using temperature data from New York City apartments, we calibrate the model and confirm its analytical predictions. Our findings provide insights into empirical demand patterns and offer actionable guidance for designing more effective DR programs and pricing policies.

SA34

Building B Level 2 B208

Global Optimization Methods and Applications

Invited Session

OPT: Global Optimization

Chair: Sergiy Butenko, Texas A&M University, College Station, TX, United States

1 - Continuous Formulations for Combinatorial Optimization Problems Based on the Probabilistic Method

Miltiades Pardalos, Texas A&M University, College Station, TX, United States, Sergiy Butenko

We present a unified framework of exact continuous formulations for a wide range of classical graph optimization problems—minimum (double) dominating set, vertex cover, maximum independent set, clique, s-plex, and max-cut—by leveraging the probabilistic method. The approach yields multi-linear polynomial programs over the unit hypercube whose optimal values coincide with the corresponding combinatorial invariants, and introduces several novel weighted extensions. To assess practical performance, we solve these formulations on numerous networks using a modern global solver (Gurobi) and five state-of-the-art nonlinear solvers (CONOPT, IPOPT, KNITRO, LOQO, SNOPT). We find that high-degree multi-linear objectives can severely impede convergence, but that systematic degree reduction—via Boole's inequality—drastically improves solution quality and robustness. The obtained results are promising and encourage a more detailed investigation of continuous formulations based on the probabilistic method in the context of solving larger-scale discrete optimization problems.

2 - New Convex Relaxations of Dynamic Systems

Kamil Khan, McMaster University, Hamilton, ON, Canada

Global optimization problems arise in several engineering applications, such as worst-case uncertainty analysis and thermodynamic modeling. Typical deterministic methods for global optimization proceed by furnishing increasingly tighter upper/lower bounds on the unknown globally optimal value. In global minimization, lower bounds are typically constructed by minimizing an appropriate convex relaxation of the original problem. Hence, improvements in convex relaxation generation lead to improvements in global optimization overall.

We present recent advances in generating useful convex relaxations for dynamic systems, with dynamics represented by an embedded ordinary differential equation model. First, we show how optimal-value functions and kinematic intuition may be deployed to yield tighter, more versatile convex relaxations of ODE solutions. Second, we show how a new rigorous finite-difference bounding method can vastly reduce the number of ODE solves required when relaxing a dynamic system, without significant sacrifices in tightness.

3 - Gpu-Based Complete and Rigorous Search for Continuous Global Optimization

Guanglu Zhang, Carnegie Mellon University, Pittsburgh, PA, United States

Continuous optimization is of significance in many science and engineering fields. In many practical applications, it is desirable, and sometimes indispensable, to find the global optimum when solving continuous nonconvex optimization problems. However, popular gradient-based methods and heuristic methods often become trapped in local optima and do not take rounding errors into account, leading to inconsistent or incorrect assumptions and applications. Based on the research gap, we introduce a GPU-based, complete, and rigorous method that can efficiently enclose the guaranteed global optimum within a user-specified tolerance for continuous optimization problems, especially for large-scale continuous nonconvex optimization problems. Using interval analysis, coupled with the computational power and architecture of GPU, the method iteratively rules out the regions in the search domain where the global optimum cannot exist and leaves a finite set of regions where the global optimum must exist. The method is validated by enclosing the global optimum of 10 multimodal benchmark test functions with scalable dimensions. These test functions represent grand challenges of global optimization, and enclosing the guaranteed global optimum of these test functions with more than 80 variables has not been reported in the literature. Our method successfully encloses the global optimum for each of these 10 test functions with up to 10,000 variables using only one GPU in a reasonable computation time, demonstrating an approximately quadratic increase in computation time with the number of variables in these optimization problems. The method is also applied to solve several practical engineering optimization problems effectively and efficiently.

4 - Symmetry-Breaking and Bilinearity Reduction in the Global Optimization of the Multiperiod Blending Problem

Daniel Ovalle Varela, Carnegie Mellon University, Pittsburgh, PA, United States, Arsh Bhatia, Carl Laird, Ignacio Grossmann

The multiperiod blending problem (MPBP) generalizes the classical pooling problem by incorporating scheduling decisions across multiple time periods, requiring blending pools to operate dynamically rather than in steady state. While various formulations and solution methods have been proposed, existing decomposition approaches often suffer from inefficiencies due to symmetric solutions. In this work, we propose an enhanced decomposition method based on Generalized Disjunctive Programming (GDP) that resolves this symmetry issue. By eliminating redundant symmetric configurations, our approach improves computational efficiency and reduces the number of bilinear constraints in subproblems by approximately 90%, resulting in subproblem solve times that are up to two orders of magnitude faster. Computational experiments show that the proposed method consistently outperforms both existing GDP-based decomposition algorithms and standard monolithic formulations solved with state-of-the-art solvers like Gurobi. To support further research, we also introduce a publicly available library of 60 MPBP benchmark instances in JSON format, designed to provide a diverse and challenging testbed for future studies.

5 - Rotational Relaxations for Global Optimization of Quadratic Programs

Yehor Blokhin, KTH Royal Institute of Technology, Stockholm, Sweden, Jan Kronqvist, Martin ryner

In this talk, we investigate the global optimization of non-convex quadratic functions subject to linear constraints defining a bounded set. We introduce a novel family of relaxation methods obtained by linearly transforming the vector space using the constraints. Preliminary experiments indicate that the presented approaches, integrated with the α BB method and McCormick envelopes, produce significantly tighter convex bounds.

SA35

Building B Level 2 B209

Fair and Dynamic Operations of Infrastructure Systems

Invited Session

Public Sector OR

Chair: Violet (Xinying) Chen, Stevens Institute of Technology, Hoboken, NJ, United States

1 - A Fairness Framework for Shared Electric Micromobility Systems: Conceptual Foundation and Case Studies**Tara Zibandehkhooy, Stevens Institute of Technology, Hoboken, NJ, United States, Violet (Xinying) Chen**

Shared Electric Micro-Mobility Systems (SEMS) are reshaping urban transportation by offering sustainable and convenient travel options. However, concerns about equitable access and usage have gained attention among policymakers and researchers. This study proposes a flexible framework for evaluating fairness in SEMS from the perspective of central decision-makers. Fairness is defined as the desirability of utility distributions—benefits or harms—across geographic or demographic stakeholder groups. The framework categorizes utilities into two types: *decision outcomes*, which capture static factors such as system supply and spatial context; and *system outcomes*, which reflect dynamic performance indicators such as vehicle availability, usage levels, and idle time. We demonstrate the applicability of this framework through two real-world case studies: a docked bike-sharing system and a dockless e-scooter system. Using operational SEMS and public infrastructure data, we compute relevant utility metrics and apply fairness measures to assess spatial equity. These case studies illustrate how the framework enables descriptive fairness analysis and supports data-driven insights for more equitable micro-mobility planning. The proposed framework serves as a foundation for systematic fairness evaluation across different SEMS types and urban contexts.

2 - Algorithmic Monoculture and Operational Hedging**Peter Zhang, Carnegie Mellon University, Pittsburgh, PA, United States, Hao Hao**

Problem. Large-model pipelines introduce a hidden single point of failure: predictions made by architecturally similar models tend to err in the same way, creating systemic risks that traditional i.i.d. assumptions or mean-squared-error metrics fail to capture.

Approach. We introduce *Algorithmic Concentration Risk* (ACR) to measure upstream prediction error alignment, and *Decision Concentration Risk* (DCR) to quantify the downstream, worst-case impact on decision-making. DCR is formulated as a distributionally robust optimization problem with a moment-based ambiguity set. To address this, we propose **RADIO** (Risk-And-Diversification Integrated Optimizer), which jointly minimizes the cost of deploying heterogeneous models and the DCR.

Novel Results.

- (i) This single, lightweight framework unifies three classical models from distinct domains: it reduces to the softmax/logit choice model under linear loss and Gumbel uncertainty; to the newsvendor model with piecewise-linear penalties; and to mean-variance (Markowitz) optimization under distributional ambiguity.
- (ii) In both heavy-tailed simulations and the public M5 retail dataset, RADIO reveals a novel “folded scaling-law,” showing how decision risk falls as investment in both accuracy and model diversity increases.
- (iii) This is the first decision-theoretic toolkit that quantifies the risk of model monoculture. It provides actionable guidance for hedging downstream decisions and pricing the value of added diversity and accuracy in predictive models.

3 - A Multi-Stage Dynamic Optimization Framework for Effective Epidemic Resource Allocation and Sustained Public Trust**Kimiya MohammadiJozani, Virginia Tech, Blacksburg, VA, United States, Sait Tunc, Esra Buyuktahtakin Toy**

Epidemics present urgent and complex decision-making challenges, where limited resources must be allocated under evolving and uncertain conditions. The stochastic nature of infection rates, driven by behavioral, environmental, and policy-related factors, adds significant complexity to planning effective interventions. Moreover, resource allocation strategies must consider efficiency and fairness to reduce the burden of epidemics while maintaining public trust. This work introduces an optimization-based framework for epidemic response, integrating disease dynamics with logistics planning under uncertainty. We formulate and solve stochastic epidemic-logistics models that enable equitable and effective allocation of critical resources, such as vaccines, across spatially and behaviorally heterogeneous populations over time.

4 - End-to-End Fairness Optimization with Fair Decision Focused Learning**Yu Wang, Stevens Institute of Technology, Hoboken, NJ, United States, Violet (Xinying) Chen**

We introduce Fair Decision-Focused Learning (FDFL), a principled framework for jointly optimizing predictions and downstream decisions under fairness constraints. Unlike traditional pipelines that treat fairness and decision-making as disjoint steps, FDFL integrates them into a unified, end-to-end learning objective. Our framework supports a broad class of fairness criteria, with a focus on differentiable decision-level objectives. We develop gradient-based training algorithms, and for a specific class of fairness measures—namely α -fairness with linear utilities—we derive closed-form optimization solutions that allow for exact gradient computation. This analytic approach significantly improves both computational efficiency and training stability.

Through healthcare resource allocation experiments, we demonstrate that FDFL yields lower decision regret and reduced prediction disparity compared to standard two-stage baselines. Moreover, using exact gradients via our closed-form solution yields better performance than approximate, perturbation-based methods.

Our results underscore the importance of aligning fairness objectives across the prediction-to-decision pipeline. FDFL offers a practical and scalable blueprint for building machine learning systems that are both effective and equitable, opening pathways for broader adoption in high-stakes domains.

5 - Tradable Equity Credits (Tecs) for Public Transit: a Dynamic Pricing Approach for Heterogeneous Travelers**Chenfeng Xiong, Villanova University, Villanova, PA, United States**

Public transit systems are vital for providing equitable access to jobs, education, healthcare, and other essential services. However, transit agencies face increasing challenges balancing financial sustainability with transportation equity, particularly in the wake of the COVID-19 pandemic. This paper introduces the Tradable Equity Credit (TEC) scheme, a novel market-based approach designed to incentivize transit agencies to enhance transportation equity while managing network demand and maximizing farebox revenue. The TEC scheme operates under a baseline-and-credit system, where transit agencies earn credits by surpassing equity targets and incur debts when they fall short. These credits are assigned monetary value and traded in an open market, providing financial incentives for agencies to adopt equity-focused policies without compromising revenue objectives. Unlike traditional subsidy programs, TEC dynamically allocates resources based on measurable equity improvements and promotes inter-agency collaboration. This paper develops a nonlinear programming (NLP) model to describe transit agencies' decision-making in the TEC market, integrating equity considerations into revenue management strategies considering heterogeneous travel demand. The model's equilibrium credit price is derived through Lagrangian dualization of the equity constraints and explored via variational inequality (VI) formulations. We propose a backpropagation-based iterative algorithm to solve the model efficiently. Case studies, including a real-world application in Sioux Falls, South Dakota, demonstrate the feasibility and effectiveness of the TEC framework. The findings show that TEC can serve as a revenue-neutral and scalable mechanism for improving transportation equity in public transit systems.

SA36

Building B Level 2 B210

Recent Developments in Bundle-Type Methods

Invited Session

OPT: Nonlinear Optimization

Chair: Jiaming Liang, University of Rochester, ATLANTA, 14620, United States

Co-Chair: Jimmy Zhang, 101 Main St, 47901

1 - A Modified Proximal Bundle Method Under Frank-Wolfe Perspective**Xu Sun, Massachusetts Institute of Technology, Cambridge, MA, United States, David Fersztand**

The proximal bundle method (PBM) is a fundamental and computationally effective algorithm for solving optimization problems with nonsmooth components. In this talk, we will present a theoretical investigation of a modified proximal bundle method, which we call the Modified Proximal Bundle with Fixed Absolute Accuracy (MPB-FA). MPB-FA modifies PBM in two key aspects. Firstly, the null-step test of MPB-FA is based on an absolute accuracy criterion, and the accuracy is fixed over iterations, while the standard PBM uses a relative accuracy in the null-step test, which changes with iterations. Secondly, the proximal parameter in MPB-FA is also fixed over iterations, while it is permitted to change in the standard PBM. These modifications allow us to interpret a sequence of null steps of MPB-FA as a Frank-Wolfe algorithm on the Moreau envelope of the dual problem. In light of this correspondence, we first extend the linear convergence of Kelley's method on convex piecewise linear functions from the positive homogeneous to the general case. Building on this result, we propose a novel complexity analysis of MPB-FA and derive an $O(\epsilon^{-4/5})$ iteration complexity, improving upon the best known $O(\epsilon^{-2})$ guarantee on a related variant of PBM. It is worth-noting that the best known complexity bound for the classic PBM is $O(\epsilon^{-3})$. Our approach also reveals new insights on bundle management.

2 - Primal-Dual Proximal Bundle and Conditional Gradient Methods for Convex Problems**Jiaming Liang, University of Rochester, Rochester, NY, United States**

This talk presents the primal-dual convergence and iteration-complexity of proximal bundle methods for solving nonsmooth problems with convex structures. More specifically, we develop a family of primal-dual proximal bundle methods for solving convex nonsmooth composite optimization problems and establish the iteration-complexity in terms of a primal-dual gap. We also propose a class of proximal bundle methods for solving convex-concave nonsmooth composite saddle-point problems and establish the iteration-complexity to find an approximate saddle-point. This talk places special emphasis on the primal-dual perspective of the proximal bundle method. In particular, we discover an interesting duality between the conditional gradient method and the cutting-plane scheme used within the proximal bundle method. Leveraging this duality, we further develop novel variants of both the conditional gradient method and the cutting-plane scheme.

3 - Nonopt: Nonconvex, Nonsmooth Optimizer**Frank E. Curtis, Lehigh University, Bethlehem, PA, United States, Lara Zebiane**

I will present my work with Lara Zebiane on an open-source C++ software package for solving locally Lipschitz minimization problems, primarily those involving objective functions that may be nonconvex and/or nonsmooth. The algorithms that are implemented in the software follow gradient-sampling and proximal-bundle methodologies. Quasi-Newton strategies with self-correcting properties play a central role in the algorithms' practical performance. In addition, a newly implemented interior-point subproblem solver greatly improves the scalability of software. I will discuss the various features of the software and present the results of numerical experiments on a variety of test problems.

4 - Linearly Convergent Algorithms for Nonsmooth Problems with Unknown Smooth Pieces**Jimmy Zhang, Purdue University, West Lafayette, IN, United States**

We propose efficient algorithms for optimizing piecewise smooth (PWS) functions, whose domain admits a finite (but unknown) partition into smooth pieces. When the objective function satisfies the quadratic growth condition, we propose a bundle level (BL) type [23] method to achieve global linear convergence. This appears to be the first global linear convergence result for any algorithm in the literature. We extend the method to handle approximately PWS functions and to solve weakly convex PWS problems, improving the current upper complexity bound by orders of magnitude to match that of smooth non-convex optimization. Moreover, we also propose the first verifiable and accurate termination criterion for optimizing the PWS function. Similar to the gradient norm for smooth optimization, it offers a tight characterization of the optimality gap under the strongly convex setting, and could be evaluated without the input of any problem parameters. A search subroutine for the certificate is developed, and we use it inside a guess-and-check scheme to improve the BL method to be almost parameter-free under both the strongly convex and the weakly convex settings.

SA37

Building B Level 2 B211

Advances in Large-Scale Stochastic and Robust Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Anton Malandii, Stony Brook University, Stony Brook, 11790, United States

1 - Nonparametric Distributional Regression via Quantile Regression**Stanislav Uryasev, Stony Brook University, Stony Brook, NY, United States**

This paper estimates the distribution of a response variable conditioned on observing factors (features). We estimate the conditional quantile of the distribution as a mixture (weighted sum) of basis quantile functions with weights depending on factors. The suggested factor model has a closed-form expression. The calibration problem is reduced to conducting quantile regressions for all confidence levels simultaneously. However, the model does not suffer from "quantile crossing" by design. The calibration is equivalent to the minimization of Continuous Probability Ranked Score (CRPS). We prove asymptotic normality of the estimator. The approach allows for the application of neural networks for the calibration of the model. Numerical experiments demonstrated the high efficiency of the approach. In particular, we have used the factor model of mixtures for predicting horse-racing outcomes.

2 - Mean-Sensitivity Tradeoffs in Distributionally Robust Optimization**Jun-ya Gotoh, Chuo university, Bunkyo-ku, Tokyo, Japan, Michael Kim, Andrew Lim**

Distributionally Robust Optimization (DRO) is a worst-case approach to decision making when there is model uncertainty. Though formulated as a single-objective problem, we show that it is intrinsically multi-objective in that DRO solutions map out a near-Pareto-optimal frontier between expected cost and a robustness measure called worst-case sensitivity (WCS). We derive explicit expressions for WCS under common uncertainty sets, enabling a principled selection of uncertainty sets based on sensitivity control. More generally, this perspective provides a quantitative measure of robustness and a sensitivity-based approach to addressing important conceptual gaps in DRO—how to choose the family and size of uncertainty sets, and how this affects the solution.

3 - Proxstorm: a Stochastic Trust Region Algorithm for Nonsmooth Optimization**Aurya Javeed, Sandia National Laboratories, Albuquerque, NM, United States**

This talk is about minimizing a smooth term plus a convex nonsmooth term. We present a stochastic proximal Newton trust region algorithm that assumes models and estimates of the objective are sufficiently accurate, sufficiently often. Like STORM (work on stochastic optimization with random models), we use facts about martingales to prove our algorithm is globally convergent with probability one.

4 - Censored Feedback Learning for Stochastic Optimization**Zhengjia Zhuo, University of Michigan Ann Arbor, Ann Arbor, MI, United States**

Stochastic optimization traditionally assumes full knowledge of the underlying distributions, but in many real-world applications, these distributions are unknown and must be learned. We use a learning framework to address this shortcoming, where in each round the algorithm implements a policy and only observes samples from the selected random variables. In addition, we consider two practically motivated feedback models: censored feedback, where only values below the chosen thresholds are observed, and binary feedback, which indicates whether/not each realization exceeds its threshold. For both settings, we obtain a near-optimal regret bound as long as the stochastic problem satisfies a natural "monotone" property. Applications of our framework include fundamental problems such as single resource revenue management problem, prophet inequality and posted pricing.

5 - Adaptive Smoothing An Importance Sampling for Stochastic Optimization of the Conditional Value-at-Risk

Anton Malandii, Stony Brook University, Stony Brook, NY, United States

We present a novel method for solving conditional value-at-risk (CVaR) optimization problems based on the dual representation of CVaR. The algorithm is based on a Bregman proximal point algorithm and consists of alternating stochastic primal and dual stages. Every (inner) primal stage involves a subproblem solved by sampling from a probability distribution updated at each dual stage (outer iteration). The dual probability distributions relative to the distribution underlying the original problem converge to the risk identifier of the solution's CVaR. Thus, the dual probabilities provide the algorithm with a built-in importance sampling mechanism that draws from the tail of the underlying distribution. Because only samples in the tail influence the CVaR, and samples outside the tail are drawn with decreasing probability, the algorithm delivers exceptional performance over other stochastic approximation methods. We prove the algorithm's convergence for convex objective functions and present numerical evidence for solving non-convex problems.

SA38

Building B Level 2 B212

Theory and Applications of Machine Learning and Optimization

Contributed Session

Contributed

Chair: Bhaskar Ray, North Carolina State University, Raleigh, NC, United States

1 - A New Mathematical Time Series Model: opt-ARIMAX

Yi-Hsuan Chu, Department of Information Management, National Chi Nan University, Nantou, Taiwan, Jing-Rung Yu, Chun-Yen Tsai, Donald Lien

Forecasting non-stationary and highly volatile financial data is critical for real-time trading and risk management, where delayed or inaccurate predictions can lead to substantial losses. The optimization-based autoregressive integrated moving average (Opt-ARIMA) method proposed by Bichescu and Polak (2023) uses a Haar wavelet transform and mathematical programming to isolate and retain white noise and automate lag terms. However, the forecasts tend to lag behind the actual data, limiting its ability to track trends accurately. To cope with this weakness, we propose an enhanced model called Opt-ARIMAX that integrates exogenous variables into the Opt-ARIMA model. This extension retains the core structure of ARIMA while improving responsiveness to real-time market changes. The Opt-ARIMAX model combines three key components: (i) using synchronous exogenous variables to capture rapid market shifts, (ii) incorporating forecast directional accuracy into the performance assessment offers a more interpretable metric, beyond conventional error measurements. Empirical results using 30-minute returns of the NASDAQ 100 indices and S&P 500 indices and corresponding index futures demonstrate that the Opt-ARIMAX offers a mathematically dynamic regression and interpretable forecasting approach. It consistently outperforms conventional ARIMA benchmarks, particularly during abrupt market regime shifts, while significantly enhancing predictive accuracy.

2 - From Vision to Optimization: AI Applications at Cargill

Madhushini Narayana Prasad, Cargill R & D, Houston/Minneapolis, TX, United States

Cargill, a global leader in food, agriculture, and industrial products, is leveraging the power of data science and artificial intelligence to drive innovation, efficiency, and sustainability across its diverse operations. In this session, we will explore how Cargill integrates cutting-edge AI technologies to solve complex business challenges and create value across the supply chain.

The presentation will begin with an overview of Cargill's mission and global footprint, followed by a deep dive into our data science strategy. We will showcase a series of project snippets that highlight the breadth and depth of our AI initiatives—from computer vision models that enhance food quality and safety, to time-series forecasting that optimizes supply chain logistics, to generative AI applications that streamline knowledge management and customer engagement, and mathematical optimization techniques that improve resource allocation and operational planning.

Attendees will gain insights into the practical deployment of AI in a global enterprise, the challenges of scaling data science solutions, and the collaborative culture that fuels innovation at Cargill. This session is ideal for data scientists, AI practitioners, and business leaders interested in real-world applications of AI in the food and agriculture industry.

3 - Dynamic Multi-Objective Optimization for Matching on a Crowdsourcing Platform

Weihang Ren, Walmart, Sunnyvale, CA, United States, Ou Sun, Hengchao Wang, Shanglin Zhou, Thomas Stiefel, Lijie Wan, Mingang Fu, Samuel Eldersveld

In this paper, we present a practical two-stage matching system designed to optimize driver-trip assignments on a crowdsourcing platform, addressing key operational challenges in real-time, multi-objective matching. Our system leverages a weighted-sum optimization approach that integrates driver preferences and efficiency constraints such as driver time-to-store to balance conflicting objectives. The platform's matching mechanism employs a flexible point-based scoring model to standardize and scale diverse metrics, enabling fair and effective comparison across objectives. Our findings highlight the system's adaptability to changing market conditions, resulting in enhanced operational efficiency and increased driver satisfaction. This applied approach offers valuable insights for scalable and adaptive matching algorithms in crowdsourced and on-demand service platforms. A switchback experiment was designed and implemented across diverse markets, revealing that our new matching system improved trip acceptance rates from 65.24% to 67.77% and matching effectiveness from 16.54% to 17.08%.

4 - Optimal Compensation Scheme when Salesperson's Environmental Knowledge is Limited: a Reinforcement Learning Simulation

Nanba Toshihiko, Kyoto University of Advanced Science, Kobe, Japan

This study revisits how firms should design sales incentives when launching a new product. Classic agency theory presumes that both the principal and the agent share, at least probabilistically, a common understanding of how effort converts into performance. At the outset of a new-product introduction, however, sales records are scant, and salespeople typically lack reliable knowledge of how their own effort will translate into revenue. We examine this information gap by embedding a reinforcement-learning algorithm within a single-period principal-agent framework. Concretely, we integrate a multi-armed bandit into the contract model and allow the salesperson to engage in soft-max exploration, progressively inferring the unknown effort-to-outcome mapping from experience.

The analysis delivers two main insights. First, relative to established products—where the effort–outcome relationship is already understood—it is profit-maximising for the firm to offer a higher commission rate for new products. Second, as learning proceeds and the salesperson refines their understanding of the underlying mapping, the optimal commission gradually converges toward the level used for existing products. These results provide a theoretical foundation for the widely observed managerial practice of offering generous launch commissions and then tapering them as market experience accumulates. The framework thereby links behavioural learning dynamics with optimal incentive design and enriches the toolbox for managing sales forces in uncertain, innovation-driven environments.

5 - Convex Clustering with Generalized Moreau Envelope Penalty

Zheming Gao, University of North Carolina Charlotte, Charlotte, NC, United States, ERIC CHI, Gal Mishne, Aaron Molstad, Chester Holtz

Convex clustering has attracted significant attention in unsupervised learning. In this research, we propose a novel convex nonconvex clustering model incorporating a generalized Moreau envelope (GME) penalty. Despite the nonconvexity introduced by the GME penalty, we establish an efficient parameter selection condition that ensures the overall convexity of the model. To solve the proposed formulation, we develop a fast forward-backward splitting algorithm. Numerical experiments on synthetic and benchmark datasets demonstrate the models effectiveness, showing advantages in clustering paths and key clustering metrics, including Adjusted Rand Index (ARI) and Variation of Information (VI). The proposed model provides reasonable and dominant clustering results in breast cancer subtype detection with TCGA data.

6 - Learning-at-Scale via RL-Enhanced Transformers and Decomposition

Seung Jin Choi, Virginia Tech, Blacksburg, VA, United States, Kimiya MohammadiJozani, Josh Cooper, Esra Buyuktahtakin Toy

This study introduces a novel Transformer-based deep learning framework for solving large-scale stochastic mixed-integer programs, with a focus on the Multi-item Capacitated Lot-Sizing Problem (MCLSP) under demand uncertainty. Our approach integrates Transformer architectures to predict optimal binary production decisions via supervised learning, followed by reinforcement learning (RL) fine-tuning that reduces infeasibility by penalizing constraint violations. To address scalability, we embed the Transformer within a decomposition framework that partitions the Single-item CLSP problem under uncertainty into a master problem and scenario-based subproblems. The model leverages Transformer parallelism to solve subproblems efficiently across multiple demand scenarios, while iterative cutting-plane generation maintains solution quality and feasibility. When integrated with the Gurobi solver, this hybrid approach significantly accelerates computation Gurobi alone, achieving near-zero optimality gaps and infeasibility rates post-RL tuning. Extensive computational experiments demonstrate that our Transformer-enhanced decomposition outperforms traditional Benders decomposition and direct MIP solvers in both speed and scalability. This work underscores the potential of combining deep learning with decomposition and cutting-plane techniques to solve complex stochastic combinatorial optimization problems at scale.

7 - L₁-Norm Regularized L₁-Norm Best-Fit Lines

Xiao Ling, Auburn University at Montgomery, Montgomery, AL, United States, James Brooks

We propose an optimization framework for estimating a sparse outlier-insensitive one-dimensional subspace. Our objective is to minimize both the representation error and a penalty on the loadings using an L₁-norm criterion for each. To our knowledge, computational methods for this pure L₁-norm formulation have not been previously proposed. Given that the problem is NP-hard, we introduce a linear relaxation-based approach. We present a novel fitting procedure, utilizing sortings of simple ratios. The proposed algorithm has a worst-case time complexity of $O(m^2 n \log n)$. We show that under certain circumstances, the method achieves global optimality. Compared to previously-proposed methods, the proposed algorithm often finds the subspace with the lowest discordance, and offers a smoother trade-off between sparsity and fit. We demonstrate the scalability with a parallel GPU implementation that provides a 16-fold improvement in computational speed for matrices of 2000x2000 over a serial CPU implementation. Furthermore, this method is distinguished by several advantages, including the lack of a need for initialization, and deterministic and replicable steps. An application to human microbiome data demonstrates the effectiveness of the algorithm in deriving meaningful and sparse solutions.

8 - Pace: Privacy Aware Collaborative Estimation for Heterogenous Glms

Bhaskar Ray, North Carolina State University, Raleigh, NC, United States, Srijan Sengupta, Aritra Mitra

With sensitive data collected across various sites, restrictions on data sharing can hinder statistical estimation and inference. The seminal paper on Federated Learning proposed Federated Averaging (FedAvg) to perform Maximum Likelihood estimation. However, FedAvg and other algorithms for parameter estimation can lead to erroneous estimation or fail to converge under model heterogeneity across sites. We propose a novel method of parameter estimation for a broad class of Generalized Linear Models with clusters of sites obtaining data based on the same distribution with possibly different values of the true parameters across clusters. It accounts for the uncertainty in the local ML estimator and that in the optimization algorithm iterates and leverages established concentration inequalities to provide non-asymptotic risk bounds. We conduct a hypothesis test-type classification based on one-shot estimation and utilize the inference to conduct a decentralized collaborative estimation, improving upon local estimation with high probability. We also prove asymptotic accuracy of the clustering algorithm and the consistency of the estimates. We validate our results with simulation studies.

SA39

Building B Level 2 B213

Advances in Mixed-Integer Optimization for Power Systems

Invited Session

OPT: Integer and Discrete Optimization

Chair: Harsha Nagarajan, Los Alamos National Laboratory, Los Alamos, NM, United States

1 - Two-Stage Robust Optimization with Decision Dependent Uncertainty for Electric Vehicle Network Expansion

Haoxiang Yang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of

Electric vehicles (EVs) have taken over 20% of the new car sales globally. The rapid development of EVs requires expanding new charging stations, which would have a major impact on the operations of power systems. Establishing new charging stations affects EV users' behavior and determines the uncertainty of the charging demand. This talk discusses a two-stage robust optimization model for EV charging station expansion. The first stage decides the location and capacity of EV charging stations. We assume the random EV charging demand lies within a polyhedral uncertainty set, which depends on the first-stage decisions. Once the uncertain charging demand is realized, the second stage optimizes the power system operations with an optimal power flow problem. We show that such a two-stage robust optimization model can be reformulated as a mixed-integer linear program and propose a column-and-cut generation algorithm to solve this problem.

2 - Enforcing Binary Decisions in Time-Domain Optimal Switching of Power Grids

Amritanshu Pandey, University of Vermont, Burlington, VT, United States

Time-domain optimization problems are crucial for addressing operational challenges in power systems, particularly if the decisions are based on the grid's dynamic states. An important subset of these problems includes optimal switching of grid equipment in and out of service while satisfying dynamic AC network constraints. These problems include binary variables representing discrete decisions, such as equipment switching and protection operations, combined with nonlinear terms from grid physics, rendering them NP-hard mixed-integer nonlinear programs (MINLP). One such MINLP strategically allocates load-shedding locations during underfrequency load shedding (UFLS) events. Commercial optimization tools fail to handle such large-scale MINLP problems robustly, especially when the problem involves thousands of binary decision variables. We present a novel approach that reformulates the underlying MINLP problem into a nonlinear programming problem (NLP) by relaxing the binary variables. To enforce binary behavior within NLP, we incorporate a penalty term, transforming the problem into a concave minimization problem with multiple local minima, each potentially corresponding to feasible binary solutions, complicating the optimization trajectory. Therefore, initially, we add a weighted smoothing function to transform the original concave problem into near-convex optimization. Subsequently, following a homotopy approach, we solve a series of subproblems systematically canceling the smoothing term while incrementally introducing the penalty term to drive toward an integer feasible solution. We demonstrate the algorithm's efficacy by solving a time-domain optimization on power system dynamic models with around 6k+ binary decision variables and 180k+ non-linear equations, to select optimal load-shedding locations to arrest frequency decline during an underfrequency event.

3 - Tightening Quadratic Convex Relaxations for the Ac Optimal Transmission Switching Problem

Harsha Nagarajan, Los Alamos National Laboratory, Los Alamos, NM, United States, Cheng Guo, Merve Bodur

The Alternating Current Optimal Transmission Switching (ACOTS) problem enhances the AC Optimal Power Flow framework by incorporating line-switching decisions, enabling cost reductions and improved system reliability. However, ACOTS involves discrete variables and nonlinear, non-convex constraints, making it computationally challenging. In this work, we propose the tightest quadratic convex relaxation for ACOTS to date by introducing several new valid inequalities, including a novel class of on/off cycle-based polynomial constraints that leverage network structure. We linearize on/off trilinear terms using a theoretically tight extreme-point representation and embed the cycle-based constraints efficiently via disjunctive programming-based cutting planes. An optimization-based bound tightening (OBBT) procedure further tightens the relaxation. To accelerate solution times, we introduce a heuristic based on a maximum spanning tree that fixes a subset of lines to remain active. Extensive

experiments on medium-scale PGLib instances show substantial improvements in relaxation bounds, while large-scale tests (up to 2,312 buses) demonstrate significant performance gains. To our knowledge, this is the first ACOTS relaxation framework to achieve near-optimal switching solutions on realistic, large-scale power grid models.

4 - Second-Order-Cone Formulations of Power Flow for Topology Optimization

Noah Rhodes, Los Alamos National Laboratory, Los Alamos, NM, United States, Jim Luedtke, Line Roald

Optimization problems that involve topology optimization in scenarios with large scale outages, such as post-disaster restoration or public safety power shutoff planning, are very challenging to solve. Using simple power flow representations such as DC power flow or network flow models results in low quality solutions which requires significantly higher-than-predicted load shed to become AC feasible. Recent work has shown that formulations based on the Second Order Cone (SOC) power flow formulation find very high quality solutions with low load shed, but the computational burden of these formulations remains a significant challenge. With the aim of reducing computational time while maintaining high solution quality, this work explores formulations which replace the conic constraints with a small number of linear cuts. The goal of this approach is not to find an exact power flow solution, but rather to identify good binary decisions, where the power flow can be resolved after the binary variables are fixed. We find that a simple reformulation of the Second Order Cone Optimal Power Shutoff problem can greatly improve the solution speed, but that a full linearization of the SOC voltage cone equation results in an overestimation of the amount of power that can be delivered to loads.

5 - Sparse Neural Approximations for Bilevel Adversarial Problems in Power Grids

Young-Ho Cho, The University of Texas at Austin, Austin, TX, United States, Mohamad Chehade, Hao Zhu

The adversarial worst-case minimum load shed (MLS) problem is pivotal for identifying critical power system contingencies under line failures. This challenge is represented as a bilevel optimization problem, where the upper level simulates an attacker determining line failures, and the lower level corresponds to the defender's generator redispatch operations. However, standard reformulation techniques using optimality conditions render the bilevel, mixed-integer formulation computationally prohibitive due to the exponential growth of candidate topologies and the non-convex nature of AC power flow physics. To address these challenges, we propose a data-driven approach that replaces conventional reformulation methods with a novel value function-based neural network (NN) model. By leveraging a sparsified, area-based partitioning of its architecture constructed via spectral clustering, the NN size is made to scale linearly with system size without compromising its accuracy in load-shed predictions. Embedding the trained NN, that approximates the value function, as a mixed-integer linear program (MILP) transforms the adversarial worst-case MLS problem into a single-level reformulation, and enables near-optimal solutions with significant computational gains. Notably, our approach ensures accurate solutions across different loading conditions, facilitating the rapid online recomputation of worst-case failures during operations. Numerical experiments on the IEEE 14- and 118-bus systems demonstrate the method's scalability and solution quality for large-scale contingency analysis, an average optimality gap of 5.8% and a maximum gap below 22.3% compared to conventional methods, while maintaining computation times under one minute.

SA40

Building B Level 2 B214

Transforming Service Systems with AI

Invited Session

Service Science

Chair: Weiwei Chen, Rutgers University, Piscataway, NJ, 08854, United States

Co-Chair: Xinyi Huang, University of Iowa, Iowa City, IA, United States

1 - Geopro-Net: Learning Interpretable Spatiotemporal Prediction Models Through Statistically-Guided Geo-Prototyping

Bang An, University of Akron, Akron, OH, United States, Xun Zhou, Ronilo Ragodos

The problem of forecasting spatiotemporal events such as crimes and accidents is crucial to public safety and city management. Besides accuracy, interpretability is also a key requirement for spatiotemporal forecasting models to justify the decisions. Interpretation of the spatiotemporal forecasting mechanism is, however, challenging due to the complexity of multi-source spatiotemporal features, the non-intuitive nature of spatiotemporal patterns for non-expert users, and the presence of spatial heterogeneity in the data. Currently, no existing deep learning model intrinsically interprets the complex predictive process learned from multi-source spatiotemporal features. To bridge the gap, we propose GeoPro-Net, an intrinsically interpretable spatiotemporal model for spatiotemporal event forecasting problems. GeoPro-Net introduces a novel Geo-concept convolution operation, which employs statistical tests to extract predictive patterns in the input as Geo-concepts, and condenses the Geo-concept-encoded input through interpretable channel fusion and geographic-based pooling. In addition, GeoPro-Net learns different sets of prototypes of concepts inherently, and projects them to real-world cases for interpretation. Comprehensive experiments and case studies on four real-world datasets demonstrate that GeoPro-Net provides better interpretability while still achieving competitive prediction performance compared with state-of-the-art baselines.

2 - Lonely Voices Online: Predictive Inference of Loneliness Behaviors from Social Media Using Decoder-Based LLMs

Jiyuan Yu, University of Iowa, Iowa city, IA, United States

Loneliness is deeply connected to a range of mental health conditions, including depression, anxiety, and suicidal ideation. As digital platforms become a primary space for emotional expression, social media provides a rich but underutilized source for understanding psychological distress at scale. However, existing computational approaches often rely on simple binary classification—determining whether a post reflects loneliness—while lacking interpretability and contextual reasoning.

This study introduces a novel inference pipeline that leverages decoder-based large language models (LLMs) to enhance both the detection and explanation of loneliness expressions in social media posts. Our method combines zero-shot and few-shot prompting strategies to classify user-generated content while simultaneously generating natural language rationales that explain why loneliness is inferred. This dual capability enables more transparent, interpretable, and empathetic assessments of online discourse.

By moving beyond black-box predictions, the proposed framework offers a scalable and minimally supervised solution for early identification of emotional distress. It holds potential applications in digital mental health monitoring, user experience design, and platform-level interventions. Our work demonstrates how LLMs can serve as both predictors and interpreters of complex psychological phenomena in online environments, contributing to the development of socially aware and ethically aligned AI systems.

3 - How Does XAI Improve Medical Diagnostic Effectiveness? Mechanisms and Methods

Dongqi Wang, Zhejiang University, Hangzhou, China, People's Republic of, Weihua ZHOU

This study examines how explainable artificial intelligence (XAI) improves diagnostic effectiveness in clinical decision-making. We first develop and empirically validate a theoretical framework grounded in two key mechanisms: perceived trust and mental model alignment. Specifically, we distinguish between two forms of alignment—Human-to-AI mental model alignment, and AI-to-Human mental model alignment. Through a controlled experiment in transfusion medicine, we demonstrate that both trust and alignment are necessary to improve the quality of human-AI collaboration.

Building on this insight, we propose a novel XAI method that aligns directly with the reasoning paradigm of evidence-based medicine. Rather than emphasizing abstract feature attributions, our approach retrieves and presents similar historical cases to support clinical decisions—mirroring how physicians reason through precedent and evidence. To enhance clinical relevance, the method leverages a large language model to distill patient context and identify the most comparable case from the historical database.

We term this approach the Retrieval Augmented Tree (RAT). By grounding AI predictions in case-based explanations that reflect medical reasoning practices, RAT facilitates intuitive understanding, fosters trust, and improves diagnostic performance—without requiring clinicians to engage with technical model internals. Together, our findings highlight both the mechanisms through which XAI enhances clinical decision-making and a practical, cognitively aligned method for its implementation in real-world healthcare.

4 - Leveraging Graph Neural Networks and Language Models for Personalized Occupation Recommendation

Yikang Wang, University of Iowa, Iowa City, IA, United States

Effective job matching requires aligning a candidate's knowledge, skills, abilities (KSAs), and motivations/interests (MIs) with occupational requirements. Traditional approaches relying solely on keyword matching often neglect nuanced motivational factors, leading to suboptimal career decisions. In this study, we propose an innovative recommendation framework that integrates large language models (LLMs) with graph neural networks (GNNs) to provide holistic job recommendations. Specifically, we utilize O*NET occupational descriptions—comprising both skills and interests—to generate semantic embeddings via a pretrained Transformer model. Occupations are connected based on semantic similarity to form a homogeneous occupation graph. Candidate resumes are similarly embedded into this latent space, naturally capturing their KSAs and MIs. A GNN then propagates information through this graph, refining occupation and candidate representations. Finally, personalized occupation recommendations are generated by comparing embeddings in the learned space. Preliminary results demonstrate that our framework significantly improves recommendation quality and interpretability compared to conventional keyword-based methods, highlighting the potential of combining structured occupational data with advanced embedding techniques for personalized career guidance.

SA42

Building B Level 3 B302

Resource Allocation and Exchange

Invited Session

Health Applications Society

Chair: Yichuan Ding, McGill University, Montreal, QC, Canada

Co-Chair: Yifeng Cao, University of British Columbia, Vancouver, BC, Canada

1 - Shift Effect and Its Impact in EDs: From Analytical Insights to Empirical Validation

Parvin Malekzadeh, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Dmitry Krass, Opher Baron, Simai He, Hongsong Yuan, Arik Senderovich

The “shift effect” in healthcare arises when patients require at least two interactions with the physician (service provider), and the provider's priorities shift depending on timing within their shift. A common instance occurs in emergency departments (EDs), where patients undergo an initial physician assessment (PIA), followed by diagnostic tests, and then a second reassessment (“disposition”). Physicians working fixed-length shifts often aim to complete care for all patients they assessed before their shift ends—driven by compensation incentives and continuity-of-care concerns. This leads to a prioritization of new patients early in the shift and in-system patients toward the end. However, this behavior may increase both the average and variability of patient waiting times, contributing to ED overcrowding. While anecdotal evidence for the shift effect is widespread, analytical investigations into its existence and impact have been limited. We address this gap through two approaches. First, a stylized queueing model shows that physicians seeking to maximize completed services naturally shift priorities over time, providing theoretical support for the shift effect. Second, using real ED data and our SiMLQ platform—an ML-based automated queueing simulator—we conduct counterfactual analysis to estimate its impact on key performance metrics, including Time to Physician Initial Assessment (TPIA) and total ED Length of Stay (EDLOS). Our findings reveal that the shift effect can significantly delay TPIA and increase EDLOS, affecting both average values and the 90th percentile. We conclude with some potential strategies to mitigate the impact of the shift effect.

2 - A 2-Stage Stochastic Programming Scheme for Online Facility Location Problems in Drone Delivery Systems

Kai Zhou, Tsinghua University, Beijing, China, People's Republic of, Michael Lingzhi Li, Kai Wang

This study addresses facility location problems in the context of drone delivery systems, which involves multiple stations and demand locations. Drones are deployed at selected stations to serve demand locations, with service time varying based on the distance between stations and demand locations. Each demand location has two attributes: demand level (indicating required service instances) and time-dependent service success probability (p-value). The p-values are heterogeneous across locations and time intervals, often initially unknown or uncertain, which exist in the objective of maximizing the sum of expected successful services across all agents. To handle the uncertainty in p-values, the problem is formulated as a combinatorial multi-armed bandit (CMAB) problem. Here, each agent servicing a demand location at a specific time represents a base arm, while the collective allocation strategy forms a super arm. In this paper, we develop a two-stage stochastic programming scheme to iteratively explore p-values and exploit optimal locations. To deal with the curse of dimensionality, we propose a novel grouping-based estimation algorithm that enhances computational efficiency while keeping the optimality gap below 5% generally. Experimental results demonstrate that this approach consistently outperforms baseline methods in balancing exploration-exploitation trade-offs and minimizing cumulative regrets. The framework effectively addresses dynamic environments where success probabilities require adaptive estimation through strategic sampling and allocation adjustments.

3 - Strategic Behavior and Uncertainty in Allocation Markets: Insights from Online Experiments

Moran Koren, BGU, Beer Sheva, Israel

We examine strategic behavior in allocation markets through three large-scale online experiments that isolate informational incentives from other effects. We investigate how individuals navigate tensions between private signals and collective decision-making across different market mechanisms. Our primary experiment (N=1,368) tests crowdfunding mechanisms in static contexts, varying signal accuracy (55% vs. 85%), group size (5 vs. 25), and participation thresholds (50% vs. 80%). Results confirm crowdfunding mechanisms induce distinct strategic behaviors compared to voting, with participants significantly less likely to follow private signals (odds ratio: 0.141, $p < 0.001$). We observe **risk aversion** and **mutual insurance**. Higher signal accuracy decreases risk aversion but increases mutual insurance reliance, validating theoretical predictions. However, contrary to theory, higher thresholds amplify risk aversion, suggesting cognitive constraints interact with strategic considerations unexpectedly. <Additional experiments examine competitive timing in organ allocation markets and herding versus delegation effects in sequential decision-making, providing complementary insights into strategic behavior across allocation contexts. Our findings advance theory by: (1) empirically validating pure informational effects in strategic behavior, (2) identifying systematic deviations highlighting cognitive limitations, and (3) demonstrating differential information aggregation across mechanisms. Crowdfunding shows comparable performance to voting under low signal accuracy but deteriorates significantly as signal quality improves. For practitioners, results offer insights for platform design around threshold mechanisms and information provision. The systematic relationship between mechanism parameters and strategic behavior suggests market designers must consider cognitive and psychological factors alongside economic incentives, with implications for crowdfunding platforms, organ allocation systems, and other collective decision markets.

4 - From Restless to Contextual: a Thresholding Bandit Approach to Improve Finite-Horizon Performance

Jiamin Xu, Cornell University, Ithaca, NY, United States, Ivan Nazarov, Aditya Rastogi, Africa Perianez, Kyrá Gan

Online restless bandits extend classic contextual bandits by incorporating state transitions and budget constraints, representing each agent as a Markov Decision Process (MDP). This framework is crucial for finite-horizon strategic resource allocation, optimizing limited costly interventions for long-term benefits. However, learning the underlying MDP for each agent poses a major challenge in finite-horizon settings. To facilitate learning, we reformulate the problem as a scalable budgeted thresholding contextual bandit problem. We carefully integrate the state transitions into the reward design and focus on identifying agents with action benefits exceeding a threshold. We establish the optimality of an oracle greedy solution in a simple two-state setting, and propose an algorithm that achieves minimax optimal constant regret in the online multi-state setting with heterogeneous agents and knowledge of outcomes under no intervention. We numerically show that our algorithm outperforms existing online restless bandit methods, offering significant improvements in finite-horizon performance.

5 - Solving for Welfare, Fairness, and Feasibility: Modern Computational Methods for the U.S. Organ Transplant System

Elijah Pivo, MIT, Somerville, MA, United States, Maura Hegarty, Thomas Athey, Nikolaos Trichakis, Dimitris Bertsimas

The U.S. organ transplant system has long contended with the challenge of fairly and effectively allocating an insufficient supply of transplantable organs. A national organ allocation policy defines how patients are prioritized for each donated organ. However, understanding the implications of an allocation policy, much less designing one capable of improving the system, is a formidable task. The waitlist consists of over 100,000 individuals spread across the country with compatible donor pools that vary drastically, each transplant involves complex and time-critical logistics, and policy changes may influence how patients and transplant centers judge transplant offers. This talk will focus on new analytical tools our group has been developing in partnership with UNOS and the OPTN, the main organizations that manage the national transplant system, in order to better support policymakers. These tools include clinical, logistical and behavioral models, a high-speed transplant system simulator, and multi-objective policy optimization. Together, these methods represent a major advance in the computational methods available for the design of organ allocation policy and an opportunity to improve the system for hundreds of thousands of patients.

SA43

Building B Level 3 B303

Advances in Healthcare Analytics

Invited Session

Health Applications Society

Chair: Joel Goh, NUS Business School, Singapore, Singapore

Co-Chair: Guang Cheng, National University of Singapore, Singapore, Singapore

1 - Using AI for Mental Health Treatment

Jingyuan Hu, UCLA Anderson School of Management, Los Angeles, CA, United States, Yican Liu, Junyou Chen

Can AI serve as a meaningful substitute for human psychologists, and how might it be integrated into everyday mental health care? We study these questions through a series of field experiments evaluating the use of an AI therapist and its broader behavioral and economic implications. In collaboration with a college population, we introduce an AI-based mental health tool, guide students through its use, and track their interactions and self-reported mental health over time. In the short run, we examine engagement and conversational patterns; in the longer run, we ask whether users develop emotional reliance on the AI and how it affects their help-seeking behavior.

2 - The Effect of Ambulance Response TIME on No-Show Behavior of Patients

Mohammadmahdi Tahergandomabadi, University of Texas at Dallas, Dallas, TX, United States, Behrooz Pourghannad, Guihua Wang, Nancy Wang

Ambulance response time is a critical performance metric in emergency medical services (EMS), with direct implications for patient outcomes and system reliability. While efforts often focus on improving speed, less is known about how response delays influence patient behavior during the wait. This study examines how ambulance response time affects the likelihood of a caller's no-show before the ambulance arrives. Using detailed operational data from a metropolitan EMS provider, we estimate the causal impact of response time on caller disengagement through an instrumental variable regression that accounts for endogeneity. After controlling for incident characteristics and location, we find that longer response times significantly increase the likelihood of caller no-show.

We also explore heterogeneity in delay sensitivity across patients with different characteristics. Older individuals and those located farther from emergency rooms show higher tolerance for delays, while younger patients are more likely to disengage. Notably, patients with more severe conditions are less patient to delays, exhibiting greater urgency and a higher likelihood of seeking alternatives when wait times are prolonged. These findings highlight the behavioral consequences of operational delays in EMS systems and offer implications for triage design, resource allocation, and targeted communication strategies aimed at reducing caller no-shows.

3 - The Double-Edged Effects of Extended Training Policies: Staffing Gains and Care Disruptions in Nursing Homes

Minyu Ye, University of Illinois Urbana-Champaign, Champaign, IL, United States, Yixin Iris Wang

In response to persistent nursing shortages in U.S. nursing homes, the federal government enacted a time-limited nurse aides in training (NAs) waiver policy. This regulation permitted nurse aides in training to remain in the workforce beyond standard 4-month certification deadlines. Although the policy aims to stabilize staffing level and improve care quality, we identify several unintended consequences that come from changes in staffing composition and care delivery. We find that although the policy increases the staffing levels of total nurse aides and improves behavioral oversight, as reflected by increased rating matrix and reduced use of antipsychotic medications, it also leads to a higher ratio of new NAs to certified nurse aides (CNAs), which negatively influences functional care. Specifically, the increased presence of new nurse aides in training disrupts consistent care routines, leading to greater reliance on catheters and wheelchairs, which in turn contributes to declines in residents' independent mobility. We also find that the policy induced greater CNA coworker variability, which leads to less efficient care in tasks that require teamwork, such as two-person mobility support and regular toileting assistance. Our results suggest that while staffing interventions can improve behavioral oversight, they may undermine functional care unless staffing schedules and team structures are adjusted accordingly. This study highlights the importance of anticipating both benefits and trade-offs when designing policies to address workforce shortages in long-term care settings.

4 - Fairness in Priority-Based Decisions with Imperfect Classification

Zhiqiang Zhang, The University of Chicago Booth School of Business, Chicago, IL, United States, Pengyi Shi, Amy Ward

We study priority-based admission decisions under the assumption that machine learning (ML) algorithms are used to generate priority scores that predict latent (unobserved) features. For instance, in the criminal justice setting—our original motivation for this work—ML algorithms predict an individual's risk of committing another crime, which can inform decisions about admission to rehabilitation programs during sentencing. As another example, in the healthcare context, ML algorithms predict a patient's underlying health condition, guiding ICU admission decisions. Under a priority-based policy, fairness can be defined in various, often conflicting ways: from the ML perspective, it includes notions such as statistical parity, predictive parity, and error rate balance; from the operations management (OM) perspective, it involves equitable allocation of limited resources. To investigate the trade-offs between fairness and efficiency, we impose various (potentially relaxed) fairness criteria and analyze their impact on admission decisions.

SA44

Building B Level 3 B304

Improving Access in Healthcare Systems: Incentives and Efficiency

Invited Session

MSOM: Healthcare

Chair: Houyuan Jiang, University of Cambridge, Cambridge, United Kingdom

1 - An Exploratory Study on How to Manage and Reduce Cancer Waiting Lists: A Queueing Approach

Xinran Hao, University of Cambridge, Cambridge, United Kingdom, Houyuan Jiang, Feryal Erhun

In the United Kingdom (UK), prolonged waiting lists for cancer treatment remain a significant challenge, with over 26,000 patients awaiting their first treatment as of November 2024. Only 69.4% of these patients received treatment within the targeted 62-day window, underscoring the urgency of timely intervention. Delays in cancer treatment can have severe consequences on survival rates, making effective management of patient flow essential. To address these challenges, we develop a queueing network model with dynamic customer types to analyze the cancer treatment pathway. Using fluid approximations of a stochastic queueing network, we derive closed-form solutions for the number of patients at each phase over time and in equilibrium, demonstrating the global asymptotic stability of the system. These analytical

solutions allow us to assess the impact of various strategies on patient abandonment rates and transitions from early to late cancer stages. A key insight from our analysis is the identification of critical capacity thresholds in congested systems. We reveal that adding capacity does not always shorten waiting times; below these thresholds, additional capacity can paradoxically increase waiting times, highlighting crucial implications for managing service systems. Our findings offer practical guidance for optimizing capacity expansion to reduce backlogs and improve early treatment rates for cancer patients.

2 - Information Disclosure, Performance Incentives and Quality Competition In Healthcare Markets

Zhan Pang, Purdue University, West Lafayette, IN, United States

Healthcare markets face persistent challenges stemming from imperfect quality information and complex interactions between payers, providers, and patients. Policymaker have been advocating three key policy instruments to improve quality of and access to care: (1) information disclosure (e.g., public reporting), (2) performance incentives (e.g., pay-for-performance), and (3) quality competition. Yet the interactions between these mechanisms remain poorly understood. Grounded in principal-agent theory and stochastic queuing dynamics, this study bridges this gap by analyzing how these policy instruments interplay to shape outcomes in healthcare markets, particularly hospital quality investments and patient choice. Our analysis reveals strategic provider responses to competing policy pressures, patterns in patient choice, and emergent system dynamics, and when these instruments complement or substitute each other. These results demonstrate that the policy effectiveness depends crucially on proper coordination. The findings provide practical guidance for designing policy portfolios that account for both provider quality improvement incentives and patient choice.

3 - The NHS Blended Payment Scheme: Incentive Issues and Optimal Reform

Kenan Arifoglu, UCL School of Management, London, United Kingdom, Sara Mohammadi, Bilal Gokpinar

The National Health Service (NHS) in England introduced the blended payment scheme for emergency care in 2019, as a mixed prospective payment model that combines fixed lump-sum payments under block contracts with variable activity-based payments under Payment by Results (PbR). Under this scheme, hospitals receive a fixed lump-sum payment at the start of the year based on their planned activity or readiness, and a variable payment at year-end that adjusts for differences between planned and actual activity, scaled by a variable rate but capped by break-glass provisions that limit extreme financial adjustment. We develop a game-theoretic model to analyze hospitals' treatment cost and planned activity decisions under the blended payment, considering information asymmetry between hospitals and regulator. Our analysis shows that while the blended scheme improves upon block contracts and PbR, it fails to fully align hospital incentives for cost reduction and readiness. We identify three key limitations: (i) information asymmetry makes it difficult for the regulator to set the appropriate variable rate, (ii) break-glass provisions may weaken hospital incentives for readiness, and (iii) hospitals may face long-term budget deficits. To address these issues, we propose a modified blended payment that retains the fixed payment but reforms the variable payment by: (i) setting the variable rate at 100%, (ii) adjusting payments based on deviations from average actual activity across hospitals rather than each hospital's own actual activity, and (iii) eliminating break-glass provisions. Our proposed payment model aligns hospital behavior with the social optimum and ensures a balanced long-term budget.

4 - Horizontal Consolidation in Healthcare Markets: Can Performance Incentives Preserve Access to Care?

Houyuan Jiang, University of Cambridge, Cambridge, United Kingdom, Zhan Pang, Sergei Savin

We study the effects of hospital consolidations on access to care in a competitive healthcare market in the presence of performance incentives. We consider a transition from a "pre-merger" market configuration where a payer delegates services to a market with two competing hospitals to a "horizontally consolidated" configuration under which the same entity manages two hospitals. We focus on two research questions: How should the payer set access-based performance incentives in response to a horizontal market consolidation? What are the effects of consolidation on patient access to care? We consider contracts where hospital compensation for delivering care includes a combination of fee-for-service payments and bonus components tied to the level of access to care patients receive. We derive the optimal bonus-type contracts that the payer can use to adapt to changes in market concentration, and we quantify the resulting impact on patient access to care and social welfare. In our analysis, we consider two alternative settings: the "fixed-FFS" setting, where the fee-for-service component cannot be altered, and the "flexible-FFS" setting, where both the fee-for-service and the bonus components can be adjusted in response to changes in the market composition. Our analysis shows that the bonus-type performance incentives provide payers with the ability to protect patient access to care upon hospital consolidation in a wide range of problem settings, provided that the intensity of competition between the hospitals is sufficiently strong. If, otherwise, the intensity of competition is relatively weak, horizontal consolidation may compromise access to care.

SA45

Building B Level 3 B305

Empirical Operations Management

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: RJ Nieuwoehner, Kelley School of Business, Indiana University, Bloomington, IN, United States

Co-Chair: Maria Ibanez, Kellogg School of Management at Northwestern University, Evanston, IL, United States

1 - Chief Sustainability Officers and Environmental Injustice: An Empirical Evaluation

Finn Petersen, Indiana University, Bloomington, IN, United States, Vibhuti Dhingra, Rachna Shah

Manufacturing facilities emit disproportionately high levels of toxic releases in underserved communities, exposing Black or African American populations to significant health risks, a phenomenon known as "environmental injustice." We study whether a firm governance reform, specifically, the appointment of a Chief Sustainability Officer (CSO), can reduce this disparity.

Using data on toxic releases from U.S. manufacturing facilities between 2000 and 2020, we find that facilities whose parent firms appoint a CSO reduce toxic emissions by up to 12% following the appointment. To identify causal effects, we employ a two-way fixed effects regression with matching, and conduct extensive robustness tests to rule out alternative explanations. We show that facilities in underserved communities experience greater reductions in toxic releases after CSO appointments, which reduces environmental injustice. This effect is driven, at least in part, by increased implementation of source reduction activities and the implementation of waste management processes by the treated facilities.

Our work sheds new light on how corporate governance reforms act as a control mechanism for addressing environmental injustice, an issue that has proven challenging to regulate effectively.

2 - Unveiling Operational Processes: The Impact of Real-Time Videos on Operational Outcomes

Zhanzhi Zheng, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Yuqian Xu, Bradley Staats

In the digital age, merchants are presented with an expanding range of resources to enhance operational transparency for customers. Moreover, the digital marketplace increasingly demands improved transparency due to the physical separation between merchants and customers. This paper explores a new method to enhance operational transparency for merchants on digital platforms: displaying real-time videos of their operational processes. Using a unique dataset about chain restaurants on a leading food delivery platform in China, we employ the difference-in-differences method with matching techniques for empirical analyses. Our results show that the display of real-time videos indeed improves restaurants' operational performance on the platform, including demand, customer average spending per order, and revenue. Furthermore, we find that reducing information asymmetry could be a key mechanism driving the positive effects of real-time videos. Specifically, these benefits are more pronounced for restaurants with less available information on the platform, such as those with fewer online reviews or lacking the premier label. Additionally, by leveraging advanced computer vision algorithms, we analyze the performance impact of key features in real-time videos through image mining, focusing on both the quality and content of the visual information that customers observe. Our findings indicate that both image quality and region of interest coverage play a crucial role in

enhancing the benefits of operational transparency. Specifically, we identify four key image attributes and operational areas detected in the videos, each significantly contributing to these benefits, though their impacts vary in magnitude.

3 - Improving Decision-Making in Resource Allocation: Evidence from Inpatient Admission Decisions

Harriet Jeon, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States, Song-Hee Kim, Hummy Song

Routing mechanisms are crucial for efficiently matching specialized resources to customer needs at the lowest cost. Improving allocation, then, is of first-order concern for many organizations. But when human decision-makers are tasked with allocation in a complex environment, it can be challenging to understand their decision processes and design systems that enhance their decision quality. Leveraging rich data derived from electronic medical records, we explore how, in the context of physicians making routing decisions of patients from the ED to inpatient medical units, decision-makers incorporate new information and adapt their decision-making processes.

4 - Crisis at the Core: Examining the Ripple Effects of Critical Incidents on Emergency Department Physician Productivity and Work Style

Lesley Meng, Yale School of Management, New Haven, CT, United States, Huifeng Su, Rohit Sangal, Edieal Pinker

Using the Emergency Department (ED) as a service setting, we study the impact of disruptive critical incidents (CIs) on worker productivity. Prior studies on CIs and worker performance predominantly focus on work environments with sequential task ordering and limited decision complexity. Since CIs occur daily in EDs but remain understudied from an operational perspective, we examine the impact of CIs in a setting with high degrees of multitasking and complex decision-making. We utilize four years of electronic medical records and administrative data from a large U.S. hospital and employ a difference-in-differences approach to analyze the impacts of CIs on worker behavior. Contrary to prior findings suggesting that CI exposure decreases worker productivity, our analysis reveals no significant reduction in the total number of patients signed and disposed of during shifts involving CIs. Physicians achieve this by working overtime, increasing reliance on diagnostic orders to facilitate decisions, and engaging in more parallel processing. However, patients discharged to home by CI-exposed physicians are 11.7% more likely to revisit the ED within seven days.

SA46

Building B Level 3 B306

Securing Tomorrow: AI-Driven Strategies for Risk Mitigation and Defense

Invited Session

Decision Analysis Society

Co-Chair: Ting Zan, University at Buffalo, Buffalo, NY, United States

1 - Robustness in Adversarial and Non-Adversarial Environments

Arunesh Sinha, Rutgers University, Piscataway, NJ, United States

Robustness to adversaries (adversarial) and environment (stochastic) is important for high stakes decision-making. In this talk, I will present two recent works on robustness against adversaries and environment. The first proposes a defense in Reinforcement Learning, where policies are highly susceptible to adversarial noise in observations. The challenge inherent to adversarial perturbations is that by altering the information observed by the agent, the state becomes only partially observable. Existing approaches fail to account for partial observability directly; we introduce a novel objective called Adversarial Counterfactual Error (ACoE), defined on the beliefs about the true state and balancing value optimization with robustness. To make ACoE scalable in model-free settings, we propose the theoretically grounded surrogate objective Cumulative-ACoE. Our method significantly outperforms current state-of-the-art approaches. The second work explores robustness to uncertain time horizon in Markov Chains, which leads to uncertainty about the cost incurred based on the state distribution when the system stops. This is critical when one needs to estimate costs within finite wall clock time. Given past data samples of how long a system ran, we theoretically analyze a distributional robust cost estimation task in a Wasserstein ambiguity set. Towards this, we show an equivalence between a discrete-time Markov Chain on a probability simplex and a global asymptotic stable (GAS) discrete-time linear dynamical system, allowing us to base our study on a GAS system only. We provide various polynomial time algorithms and hardness results for different cases in our study, including a fundamental result about Wasserstein distance based polytope.

2 - Robust Optimization with Diffusion Models for Green Security

Lingkai Kong, Harvard University, Boston, MA, United States

In green security, defenders must forecast adversarial behavior, such as poaching, illegal logging, and illegal fishing, to plan effective patrols. These behaviors are often highly uncertain and complex. Prior work has leveraged game theory to design robust patrol strategies to handle uncertainty, but existing adversarial behavior models primarily rely on Gaussian processes or linear models, which lack the expressiveness needed to capture intricate behavioral patterns. To address this limitation, we propose a conditional diffusion model for adversary behavior modeling, leveraging its strong distribution-fitting capabilities. To the best of our knowledge, this is the first application of diffusion models in the green security domain. Integrating diffusion models into game-theoretic optimization, however, presents new challenges, including a constrained mixed strategy space and the need to sample from an unnormalized distribution to estimate utilities. To tackle these challenges, we introduce a mixed strategy of mixed strategies and employ a twisted Sequential Monte Carlo (SMC) sampler for accurate sampling. Theoretically, our algorithm is guaranteed to converge to an epsilon equilibrium with high probability using a finite number of iterations and samples. Empirically, we evaluate our approach on both synthetic and real-world poaching datasets, demonstrating its effectiveness.

3 - A Global Multi-Echelon High-Tech Supply Chain Network Equilibrium Under Trade-War Export Sanctions

Ting Zan, University at Buffalo, Buffalo, NY, United States

Export sanctions (ESs) have become prominent trade policies amid heightened geopolitical tensions, reshaping international trade and cross-border supply chain networks (CSCNs). This study investigates how ES affects firm-level decision-making and evaluates the effectiveness of government-subsidized research and development (R&D) innovations in mitigating these impacts. We first develop a comprehensive network equilibrium model that involves multiple suppliers, manufacturers, and demand markets in both Bertrand and Cournot competition. Subsequently, we derive equilibrium conditions using variational inequality theory, establish their qualitative properties, and implement the Euler algorithm to solve the model numerically. Extensive numerical experiments based on real-world scenarios illustrate that ES impacts vary significantly according to the network structure, geographical region, supply chain layer, and market parameters. Although R&D innovation does not entirely offset supply chain disruptions, it effectively mitigates the adverse effects of ES in targeted regions. Furthermore, our findings indicate conditional misalignment in preferences regarding R&D outcomes among suppliers, manufacturers, and government entities.

SA47

Building B Level 3 B307

Governance and Regulation

Contributed Session

Contributed

Chair: Amin Rahimian, University of Pittsburgh, 3700 OHara Street, Pittsburgh, PA, 15261, United States

1 - From Knowledge to Open Innovation: Unlocking Circular Economy Potential in Smes Through Intellectual Capital and CSR

Bilal Iftikhar Makki, University of science and technology of China, Hefei, China, People's Republic of, Nabeel Rehman, feng feng

The transition from a linear "take-make-waste" approach to a circular economy is no longer an option but a need for sustainable economic development.

However, for SMEs in developing economies, this transformation is challenging. Overcoming the challenges requires a strategic approach like open innovation, a key enabler of circular economy principles for sustainable value creation. This study examines how knowledge management impacts open innovation through the sequential role of intellectual capital and corporate social responsibility. Additionally, it explores the moderating role of organizational and employee-level behavioural factors, such as resource efficiency practices and employee readiness for organizational change. Data was collected from 402 employees working in SMEs within the manufacturing sector of a developing country. The study employed partial least squares structural equation modeling (PLSSEM) to test the hypothesized relationships. The results confirm that knowledge management significantly influences intellectual capital and CSR, both of which sequentially enhance open innovation. Additionally, resource efficiency moderates the relationship between intellectual capital and open innovation but not between CSR and open innovation. Conversely, employee readiness for change moderates the CSR-open innovation relationship but does not moderate the intellectual capital-open innovation relationship. This research presents a novel contribution by combining knowledge management, intellectual capital, corporate social responsibility, and behavioral elements within a single framework, particularly regarding small and medium-sized enterprises in developing countries. It emphasizes the sequential and moderating factors that influence open innovation, providing practical recommendations for companies seeking to shift towards sustainable and resource-efficient practices

2 - Long-Term and Intra-Day Learning and Forgetting in Gig Work: Evidence from Online Grocery

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

Online platforms operating on a gig-contractor model rely on self-scheduling gig workers to meet real-time demand, while maintaining the desired productivity and service quality levels. In a gig-work setting, learning takes place primarily through task repetition and familiarity with the work environment. A factor that can deter learning in this setting is gaps in the continuity of service. Since gig workers can choose when and for how long to work, there is a high likelihood that service gaps lead to 'forgetting'. Drawing from theories of learning and forgetting, we investigate whether intra-day experience can mitigate the detrimental effects of forgetting on picking performance. Using data from an online grocery platform, we develop an econometric model to analyze productivity and service quality, based on the intra-day and long-term learning and forgetting of workers, while accounting for sample selection bias and endogeneity. Our findings reveal that as intra-day learning increases, productivity and service quality improves. We further examine the role of task batching and task complexity on performance and observe that when workers batch orders, higher intra-day learning reduces picking time but increases delays. For complex tasks, higher intra-day experience is beneficial up to a threshold, beyond which it improves productivity but reduces substitutions. Finally, we develop a task allocation algorithm that utilizes intra-day and long-term learning to allocate tasks to workers based on the nature and complexity of the tasks. We demonstrate that assigning tasks to workers based on task type, accounting for recent and long-term learning, leads to performance improvements.

3 - Does Helping Hurt? How Food Delivery Wage Protections Disrupt Ride-Sharing

Yunmin Choi, Korea University Business School, Seoul, Korea, Republic of, Kyeongsu Min, Jiye Baek

This study investigates the unintended cross-platform spillover effects of government regulations designed to protect platform workers. Specifically, we examine how the introduction of minimum hourly wage standards, which shift compensation from per-task to hourly pay for app-based food delivery workers, affects ride-sharing usage. As location-based gig platforms become increasingly interconnected, targeted regulations can disrupt labor allocation across sectors and alter the matching process between clients and workers. We argue that labor protections targeting a specific platform can produce spillover effects on adjacent platforms not directly subject to the regulation, thereby impacting the broader economy. Using a counterfactual estimation approach, we provide empirical evidence that implementing hourly pay for delivery workers reduces the availability of ride-sharing drivers, as indicated by a decline in ride-sharing trips and an increase in average fares. This study contributes to the literature on platform governance and labor market dynamics in multi-sided platforms. Our study also offers practical insights for policymakers and platform businesses facing growing pressure to improve working conditions.

4 - Short-Term Rentals and Their Dual Impact in College Towns: A Multi-Site Panel Study on Housing Affordability and Hotel Market Displacement

Aadi Kulkarni, UVA, Nashua, NH, United States

This study measures how the rapid rise of short-term rentals (STRs), such as Airbnb listings, affects both long-term housing costs and hotel performance in U.S. college towns. We merge monthly data from AirDNA, Zillow, STR Global, the Bureau of Labor Statistics, and the National Center for Education Statistics for ten college communities between January 2020 and December 2021. Using a difference-in-differences design and two-way fixed-effects panel regressions, we relate changes in STR density to city-level median rents and hotel revenue per available room (RevPAR), while controlling for employment, income, enrollment, and event weeks. Results show that adding one STR for every 1,000 housing units raises median rent by just over four percent, with the effect climbing to six percent in towns where new construction is

slow. In parallel, a one-unit increase in STR density is linked to a 1.8 percent drop in hotel RevPAR, a decline that grows during football and graduation months. These findings suggest that STR growth squeezes renters and trims hotel windfalls in college markets. Cities may therefore need locally tailored caps or licensing rules, while universities should expand student housing and open dorms during peak events to ease market pressure.

5 - Privacy-Preserving Location Sharing in Crowd-Sourced Shipping

Amin Rahimian, University of Pittsburgh, Pittsburgh, PA, United States

In applications involving volunteer delivery of food and other supplies to resource-limited locations as well as crowd-shipping of deliveries from a depot, a central dispatcher relies on location sharing from crowd-shippers to design optimal matching of deliveries to potential crowd-shippers and volunteers. However, this poses privacy concerns in sharing sensitive location and destination information between the dispatcher and shippers that can disrupt the dispatch operations if crowd-shippers are not afforded reliable privacy protections. We investigate privacy-preserving dispatch algorithms that protect destination location information shared by the shippers while maintaining the overall performance and the quality of matched deliveries. We investigate the cost of privacy in both local and central privacy regimes, where the destination information is protected either at the user-level locally or the matched deliveries are randomized centrally at the dispatch to protect the location information. We investigate managerial insights and policy implications that determine who bears the cost of added privacy protections and how the compensations to crowd shippers should vary accordingly.

SA48

Building B Level 3 B308

Data-driven Decision Making

Invited Session

Decision Analysis Society

Chair: Leann Thayaparan, Johns Hopkins University, Baltimore, MD, United States

1 - Centralized vs. Host-Set Pricing: Lessons from a Car-Sharing Marketplace

Christopher Dragomir, University of Pennsylvania, Philadelphia, PA, United States, Neha Sharma, Ioannis Stamatopoulos

Platforms now mediate most peer-to-peer transactions, yet price frictions persist. Determining who should set prices on two-sided marketplaces is unclear: platforms hold superior information on demand and supply, while hosts best understand their own costs. We explore the tradeoffs with these pricing mechanisms through data from a natural experiment wherein a large car sharing platform let hosts decide prices for their cars or accept the platform's smart pricing.

2 - Interpretable Feature Engineering Through Clustering: from Healthcare to Retail

Matea Gjika, Massachusetts Institute of Technology, Cambridge, MA, United States, Alessandro Morosini, Georgia Perakis, Marc Saouda

High dimensionality, multicollinearity, and noise in datasets can significantly hinder the interpretability and performance of predictive models. We address these challenges by introducing a clustering-based approach on the feature space to construct more meaningful features that preserve essential information while reducing granularity. Our method leverages domain knowledge to identify the appropriate clustering criteria and harnesses the power of mixed-integer optimization to make the clustering problems tractable. We present two case studies: in healthcare, we show that correlation-based gene-clustering improves the interpretability and predictive power of cancer treatment models, while providing actionable insights for personalized treatment. In retail, we present a weather state clustering framework that allows us to quantify and understand the impact of weather on sales for one of the world's largest fashion retailers, with the goal of improving their distribution decisions. Our approach leads to models that are more robust and interpretable, ultimately creating the opportunity for more effective data-driven decision making.

3 - Increasing Parents' Engagement on Ed-Tech Platforms: Evidence from the Field

Alex Akira Okuno, NYU Stern, New York, NY, United States, Divya Singhvi, Somya Singhvi

Ensuring universal access to quality education remains a formidable challenge in developing countries. A key issue affecting educational outcomes among low-income households is limited parental involvement in their children's education. EdTech platforms can help build cost-effective and impactful solutions to address this problem. In collaboration with Rocket Learning, one of India's leading non-profit EdTech organizations, we design, implement, and optimize nudges to increase parent engagement in digital educational communities. We conduct a large-scale Randomized Control Trial (RCT) with nearly 154,000 parents to investigate the effectiveness of peer comparison and self-comparison nudges in improving engagement. Our results demonstrate that nudges can significantly increase the number of engaged users by 8% to 13.5%. We find heterogeneous effects influenced by teachers' activity levels, parents' pre-treatment engagement, and potential benefits from targeted nudging. Leveraging RCT data, we develop machine learning-based targeting policies to optimize nudge assignments. A second RCT with 122,000 parents reveals that while targeted nudging performs comparably to the best uniform policy on average, it identifies a substantial sub-population (8.8% of parents) for whom targeted nudges improve engagement by 5.4% over the uniform policy. These findings have significant implications for policymakers and non-profits focused on improving educational outcomes in Low and Middle-Income Countries (LMICs), demonstrating the effectiveness of nudges in increasing parental engagement, revealing significant heterogeneity in their impact, and highlighting the nuanced benefits and challenges of ML-based targeting strategies in educational interventions.

4 - Grid-Scale Mobile Battery Energy Storage Systems

Vishrut Rana, The Wharton School, Philadelphia, PA, United States, Christian Kaps, Serguei Netessine

The rapid rise of renewable energy introduces operational challenges for electric grids due to the intermittent nature of solar and wind generation. To manage this intermittency, energy storage technologies have seen considerable interest, and among the many technologies, Battery Energy Storage Systems (BESS) have emerged the most commercially viable. BESS store excess energy when renewable supply is high, and thus price of electricity is low, and release it when demand, and thus price, spikes. Profitability of BESS depends largely on nodal electricity price volatility. However, this price volatility can diminish over time due to grid expansion or increased investment in storage and generation capacity in a region. This makes long-term investment in fixed-location BESS increasingly uncertain, especially as grid conditions evolve year over year. Mobile Energy Storage Systems (MESS) offer a promising alternative. Unlike stationary systems, MESS can be relocated between grid nodes to follow shifting price signals and better align with short-term market conditions or long-term infrastructure changes. This flexibility allows MESS to maintain financial viability while also providing grid operators with a dynamic tool to address localized reliability and congestion challenges. In this paper, we measure the financial performance of MESS using market data from the PJM Interconnection and empirically identify the scaling opportunity for MESS fleets.

5 - Tractable Approximation to Choice Models for Multiple Product Pricing

Xuening Wang, Massachusetts Institute of Technology, Cambridge, MA, United States, Georgia Perakis

In price optimization, demand models must balance behavioral interpretability with computational tractability. While discrete choice models like the Mixed Multinomial Logit (MMNL) offer behavioral realism, they are often intractable for price optimization. On the other hand, price optimization literature often assume a much simpler model that admits minimal behavioral structure for computational ease. Motivated by practical needs in retail settings, this paper proposes a class of piece-wise linear functions to approximate the demand function from discrete choice models. We introduce the Contracting Localization Update (CLU) algorithm to efficiently estimate these functions from historical transaction data. We provide analytical guarantees for our methodology and validate the performance of our approximation through numerical experiments, achieving on average over 86.9% of optimal revenue under MMNL ground truth. Our method bridges the gap between interpretable modeling and efficient optimization and is adaptable to broader choice structures, including complementarity and multi-purchase behavior.

SA49

Building B Level 3 B309

Frontiers in Supply Chain Analytics

Invited Session

MSOM: Supply Chain

Chair: Li Chen, Cornell University, Ithaca, NY, United States

Co-Chair: Ruihao Zhu, Cornell University, Ithaca, NY, United States

1 - Thompson Sampling for Repeated Newsvendor

Weizhou Zhang, National University of Singapore, Singapore, Singapore, Li Chen, Hanzhang Qin, Yunbei Xu, Ruihao Zhu

We investigate the performance of Thompson Sampling (TS) for online learning with censored feedback, focusing primarily on the classic repeated newsvendor model—a foundational framework in inventory management—and demonstrating how our techniques can be naturally extended to a broader class of problems. We model demand using a Weibull distribution and initialize TS with a Gamma prior to dynamically adjust order quantities. Our analysis establishes optimal (up to logarithmic factors) frequentist regret bounds for TS without imposing restrictive prior assumptions. More importantly, it yields novel and highly interpretable insights on how TS addresses the exploration-exploitation trade-off in the repeated newsvendor setting. Specifically, our results show that when past order quantities are sufficiently large to overcome censoring, TS accurately estimates the unknown demand parameters, leading to near-optimal ordering decisions. Conversely, when past orders are relatively small, TS automatically increases future order quantities to gather additional demand information. Extensive numerical simulations further demonstrate that TS outperforms more conservative and widely-used approaches such as online convex optimization, upper confidence bounds, and myopic Bayesian dynamic programming. This study also lays the foundation for exploring general online learning problems with censored feedback.

2 - Llm-Based Choice Prediction: a Generalized Direct Preference Optimization Approach

Jinsong Liu, Cornell University, Ithaca, NY, United States, Li Chen, Dongdong Ge, Ruihao Zhu

In this talk, we propose a LLM-based approach for customer choice prediction. We develop the Generalized Direct Preference Optimization (GDPO) method that can effectively utilize the customers' historical choice data to fine-tune a pre-trained LLM. GDPO extends the conventional DPO method by additionally incorporating the crucial tie information in the data. We not only establish a theoretical performance guarantee for our approach, but also demonstrate its excellent performance through extensive numerical experiments on both synthetic and real-world datasets.

3 - Policy Learning with Switch Estimators: Statistics, Optimization and Computation

Hanzhang Qin, National University of Singapore, Singapore, Singapore, Jingren Liu, Junyi Liu, Mabel C. Chou, Jong-Shi Pang

We study an offline policy learning problem: Given an offline data set generated by a historical policy, we are interested in (i) evaluating the counterfactual outcomes had a target policy been implemented during the time of data collection, and (ii) learning the optimal policy. We employ a sample-dependent adaptive evaluator (SDAE) that interpolates between the direct regression and the augmented inverse propensity weighting (AIPW) estimator to balance the bias-variance tradeoff in policy evaluation. Unfortunately, the SDAE-based policy learning problem is nonconvex and nondifferentiable, posing significant challenges for optimization. We reformulate the policy learning problem as a Heaviside composite optimization problem (HSCOP), which can efficiently solved by the progressive integer programming (PIP) method. Beyond the computational efficiency guarantees, we show that the SDAE achieves minimax rate optimal regret in learning the optimal policy.

4 - A Variational Inference Method for Bayesian Dynamic Control

Li Chen, Cornell University, Ithaca, NY, United States, Jiachun Li, Jue Wang

In this talk, we introduce a variational inference method for a class of Bayesian dynamic control problems involving either exact or censored observations, with common applications in operations management. We show that this method is asymptotically optimal under some mild conditions. Finite-horizon performance bounds for this method are also obtained and numerically illustrated.

5 - Fairness-Aware Static and Dynamic Assortment Optimization: Optimal Selection with Balanced Market Share

Qing Feng, Cornell University, New York, NY, United States, Omar El Housni, Huseyin Topaloglu

Assortment optimization plays an important role for many online retailers in maximizing their revenue. However, assortment decisions exclusively for revenue maximization often create significant disparity in sales among products offered by the retailer. To address the fairness concerns in assortment optimization, we introduce a market share balancing constraint to guarantee fairness among offered products, ensuring that the expected sales of any two offered products differ by at most a factor of α given exogenously. We consider both the static and dynamic assortment optimization problem under MNL with market share balancing constraint. In the static problem, the seller chooses a distribution over assortments such that the market share balancing constraint is satisfied in order to maximize the expected revenue. We show that the problem can be solved in polynomial time. Specifically, we show that a product is offered in the optimal solution if and only if both its revenue and preference weight exceed certain thresholds. As an extension, we also consider the setting where besides market share balancing constraint, there are additional constraints on the set of offered products. We also study the dynamic problem, where each product has an initial inventory, and the seller chooses a dynamic control policy to maximize the total expected revenue such that the realized sales of each product do not exceed its initial inventory, and the market share balancing constraint is satisfied in expectation. We establish an asymptotically optimal policy for the dynamic problem, whose approximation ratio converges to one as the initial inventories increase.

SA50

Building B Level 3 B310

Theory and Applications of Machine Learning for Optimization

Contributed Session

Contributed

Chair: Jing Wang, Rutgers University, 132 Yale Ave, East Newark, 07304, United States

1 - Hierarchical Risk Parity Portfolio Optimization with Constant Absolute Risk Aversion

Kwame Addey, Minnesota State University Moorhead, Moorhead, MN, United States

Diversification is the typical investment strategy of risk-averse agents. However, most portfolio optimization models often ignore the importance of investor's degree of risk aversion. This paper proposes an investment strategy based on the hierarchical risk parity (HRP) optimization model under constant absolute risk aversion (CARA). The HRP model innovatively employs graph theory and machine learning algorithms to optimally utilize hierarchically hidden information to create well-balanced and resilient investment portfolios. The results reveal better performance for HRP in annualized returns, Sharpe, Calmar, and Sortino ratios. However, the higher order risk aversion HRP outperforms the HRP in terms of volatility, tail risk-adjusted returns, and maximum drawdown. The proposed model is important for portfolio managers because it presents the leverage of choosing an investment strategy based on objectives, risk tolerance, market conditions, and investment horizon.

2 - Penalized Quantile Regression from An Optimization Perspective

Cezareo Rodriguez, SimpleRose, Saint Louis, MO, United States

In real-world data settings, especially those with skewed distributions, heteroscedasticity, or outliers, standard least squares regression often yields misleading results. Penalized quantile regression offers a robust alternative to classical least squares estimation by modeling conditional quantiles of the response variable while controlling for model complexity. In this talk we investigate LP and MIP formulations of penalized quantile regression under certain penalty functions, and how one can leverage parallelism and primal-dual relationships to create robust statistical models of entire distributions. We also present novel algorithmic strategies that exploit the structure of the optimization formulations to further enable scalable computation. Empirical results demonstrate the effectiveness of our methods on both synthetic and real-world datasets, and offer insight into how optimization-based techniques can be useful for scaling up robust regression models.

3 - Lie Theory-Based Explanation of Neural Network-Generated Quantum Gate Parameters for Barren Plateau Mitigation

Zhehao Yi, The University of Alabama in Huntsville, Huntsville, AL, United States, Rahul Bhadani

Barren plateaus describe the phenomenon in variational quantum algorithms where gradients vanish exponentially with the increase number of qubits, causing training to stall due to ineffective parameter updates. Our work offers a mathematical explanation for why neural networks, when generating quantum gate parameters, can effectively mitigate the barren plateaus problem in variational quantum algorithms. By applying tools from Lie theory, we map the single-qubit quantum gates into a geometrically structured group $SU(2)$, where the evolution of parameters can be more naturally analyzed. Our results show that the updated trajectories of neural network-generated parameters are one of the geodesics in the $SU(2)$ group, indicating a mapping between the parameter update process and the geometry of the quantum model. This finding provides a mathematical explanation for the success of neural networks in optimizing quantum circuit parameters. In particular, we reveal the mathematical principles underlying why parameter updates by neural networks significantly reduce the likelihood of gradient vanishing, which is that the optimization process is guided along the direction of geodesics to find the best parameters. This finding provides a mathematical groundwork for the class of methods that use neural networks to generate parameters mitigating barren plateaus, and also for future studies on hybrid quantum-classical architectures and their optimization procedure.

4 - Multitask Sufficient Dimension Reduction with Common Support

Jiaying Weng, Bentley University, Waltham, MA, United States

In this talk, we will focus on developing a new dimension reduction approach for multitask learning. Sufficient dimension reduction is a contemporary machine learning technique that reveals the underlying structure of high-dimensional data. Multitask learning, on the other hand, is a transformative approach that leverages shared information across multiple related tasks to enhance prediction accuracy and deepen feature understanding. Integrating sufficient dimension reduction within the multitask learning framework will markedly improve the analysis of complex, high-dimensional data across multiple tasks. We propose a multitask sufficient dimension reduction (MT-SDR) framework, starting with a multitask nonlinear regression model. We develop a novel optimization problem to estimate the central subspace for each task. We then establish the minimax lower bound of the sparse estimator for the multitask setting, generalizing the single-task result to this context. Furthermore, we show that the proposed MT-SDR estimator achieves a nearly optimal rate of convergence.

5 - Energy-Based Neural Surrogates to Guide Large-Scale Milp Solving

Niki Triantafyllou, Imperial College London, London, United Kingdom, Maria M. Papathanasiou

Solving large-scale mixed-integer linear programs (MILPs) remains a core challenge across industries due to their combinatorial complexity and high computational cost. While machine learning (ML) has shown promise in accelerating or approximating optimization, many approaches rely on directly predicting discrete decisions, which can violate feasibility or lead to poor-quality solutions when constraints are complex or tightly coupled.

We propose an energy-based learning (EBL) framework that integrates deep learning with structured optimization, enabling the efficient solution of large-scale MILPs while improving feasibility and solution quality. Rather than predicting discrete decisions directly, our approach trains neural networks to learn an energy function over the feasible space that assigns lower energy to high-quality, feasible solutions and higher energy to infeasible or suboptimal ones.

During inference, we minimize the learned energy subject to the original MILP constraints, combining the expressiveness of neural models with the rigor of exact optimization. Our method builds on earlier work using neural networks to approximate binary decisions in large-scale supply chain models, and addresses key limitations by ensuring constraint satisfaction through energy-based inference.

We apply our framework to a large-scale MILP model for investment and tactical planning in personalized medicine supply chains. Results show that the proposed approach consistently improves feasibility and objective performance compared to direct prediction methods. By embedding domain knowledge into the energy function and preserving optimization structure, our method bridges data-driven learning and combinatorial optimization, offering a scalable solution for complex decision-making problems.

6 - On Representing Optimization Problems with Graph Neural Networks

Jialin Liu, University of Central Florida, Orlando, FL, United States

The recent era has witnessed the success of machine learning models across various domains, including optimization. In these applications, ML models typically capture key features of problem instances to suggest adaptive configurations or provide effective warm-starts, thereby accelerating the solving process. This talk will focus on graph neural networks (GNNs), discussing their expressive power in approximating key mappings from problems to their features—such as feasibility and boundedness—or their solutions. We will explore what GNNs can and cannot do in this context.

7 - Integrating Neural Network Forecasts into Regional Power System Planning Models

Jing Wang, Rutgers, Jersey City, NJ, United States

This study explores a modeling framework that embeds trained neural networks (NNs) into a multi-regional generation expansion planning (GEP) model to support long-term electricity system planning. The optimization problem is formulated as a mixed-integer linear program (MILP) that jointly captures investment decisions, inter-regional power exchanges, and operational constraints under a cost-minimization objective.

A feedforward neural network with ReLU activation is incorporated into the GEP model using a piecewise-linear representation. This allows the model to utilize learned nonlinear relationships—such as demand or cost forecasts—while preserving mathematical compatibility with MILP solvers. We begin with a simplified architecture inspired by prior work, in which all network weights are nonnegative. This structural constraint reduces modeling complexity and facilitates initial testing, though it also limits the network's expressiveness.

We investigate whether embedding NN-based predictions alters key model outputs—including investment choices, regional generation mix, and total system cost—or primarily affects solution convergence. Comparative scenario analysis is conducted to assess the impact of network predictions under varying demand and transmission assumptions.

This research contributes to the broader integration of machine learning and energy systems optimization by evaluating practical approaches for combining neural forecasts with structured decision models. It also sets the stage for future work involving more expressive networks and broader classes of uncertainty representations in large-scale optimization frameworks.

SA51

Building B Level 3 B311

Empirical and Analytical Approaches in Healthcare Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Nazli Sonmez, ESMT GmbH, Berlin, Germany

1 - The Impact of Prescription Drug Monitoring Programs on the Structure of the Opioid Supply Chain

Elaheh Alimohammad, University of Oregon Lundquist College of Business, Eugene, OR, United States, Behrooz Pourghannad, Guihua Wang

The opioid epidemic has resulted in over one million deaths in the U.S. since 1999. While Prescription Drug Monitoring Programs (PDMPs) have been implemented nationwide to mitigate misuse by tracking opioid prescriptions, little is known about their impact on the underlying supply chain structure. This study utilizes a difference-in-difference design to investigate the impact of must-access PDMP mandates on the decisions and behaviors of key supply chain actors.

2 - Is Continuity All We Need? a Modeling Approach to Evaluating Relational Continuity in Primary Care

Naireet Ghosh, London Business School, London, United Kingdom, Nicos Savva, Yueyang Zhong

Relational continuity of care – the ongoing therapeutic relationship between a patient and the same physician – has been repeatedly linked to lower mortality, fewer hospitalizations, and reduced consultation demand. Yet strict continuity is operationally challenging as it constrains resource pooling among physicians, consequently lengthening queues and raising waiting times. To analyze this tradeoff, we develop a multi-server queueing model in which each patient chooses (or is routed to) either their regular physician or the earliest available physician, capturing the dynamic feedback whereby disrupted continuity leads to poorer health outcomes leading to increased visit rates in the future. We show that, from the perspective of a social planner, it is optimal to maintain intermediate levels of continuity wherein patients with the most complex needs visit their regular physician most of the time while those with lower complexity always visit the earliest available physician. This balance depends on factors such as the size of the practice, the patient mix, and the level of congestion. Leaving the decision to patients alone produces extreme Nash equilibria – either always maintain continuity or always abandon it – both of which can be far from the social optimum, highlighting the need for central coordination. We calibrate the model to data from NHS general practices and numerically analyze the optimal and equilibrium continuity policies in those contexts. The case study provides quantitative guidance for NHS administrators deciding how much continuity to prioritize at different practices and when to retain control over patients' appointment scheduling.

3 - To Each Their Own (Shifts): Incorporating Heterogeneous Worker Preferences into Shift Work Schedules

Hummy Song, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States, Harriet Jeon, Song-Hee Kim

Shifts are the dominant way to work in many contexts requiring 24/7 coverage, including call centers, police departments, and hospitals. While the detriments of shift work are well-documented both at the individual and organizational levels, its deployment is often unavoidable given round-the-clock staffing needs. We explore a potential organizational lever – incorporating preferences over shift design – to mitigate ramifications of shift work in the context of acute care bedside nurses. Using survey, administrative, and shift data, we examine whether and the extent to which individual choice over dimensions of their shifts mitigates the impact of shift work on work (dis)satisfaction and turnover of nurses.

4 - Feature-Driven Priority Queuing

Simrita Singh, Santa Clara University, Santa Clara, CA, United States

Traditional queuing theory assumes that customer types are perfectly observed, assigning each customer to a type-specific priority queue – this is known as type-driven priority queuing. We study feature-driven priority queuing, where types are not perfectly observed but inferred from features using a classifier. Our work introduces an explicit modeling of the classifier within the queuing system design, which is novel in this context. We analyze the optimal queue classification that minimizes queuing delay costs and compare a plug-and-play, type-first approach – where type probabilities are predicted by a classifier and then used to optimize queuing – with a direct approach that optimizes the classifier to predict queue probabilities directly from features.

Our primary contribution is proving that the direct approach outperforms the type-first approach, except when the type classifier perfectly predicts the Bayes posterior. To illustrate how the direct approach achieves this improvement, we use a combination of analytical results, numerical examples, and simulations on a large real-world dataset. In our numerical study using state-of-the-art image classifiers on a dataset of medical images used in radiology triage, the direct approach reduces delay costs by 30-60% relative to the type-first approach, particularly in high-utilization scenarios.

SA52

Building B Level 3 B312

Technology and Future of Work

Invited Session

MSOM: Service Operations

Chair: Qiuping Yu, Georgetown University, Washington, DC, United States

1 - (Machine) Learning Preferences from Complex Choice Sets: An Application to Service Networks

Kenneth Moon, Cornell University, New York, NY, United States

Hidden complexity pervades commonplace choices. For instance, at everyday venues such as supermarkets and amusement parks, customers routinely consider astronomically many possible choices of paths through some or all of the venues' stations (e.g., sections or rides). Rich data now capture such customer choices, but the complexity remains a challenge for empirical research. Firstly, it is unclear how customers assess such complex decisions. Secondly, choice estimation methods exhibit fundamental issues ranging from intractability to inconsistency when estimating consumer preferences from such data. Under commonplace conditions, we show that a hidden but provably low-dimensional fraction of relevant choice comparisons suffices to capture all customer preference information from data and can be automatically learned and exploited by machine learning (ML). The same low dimensionality further justifies simple shopping heuristics for customers. We develop a neural network-based estimator of customer preferences and demonstrate it on data tracking shoppers at a hypermarket serving over 1M customers annually. Such estimation uncovers the hypermarket's demand structure, and modest capacity reallocations counterfactually raise equilibrium service throughput by 5-25% through. Addressing a long-standing challenge in marketing and operations, we identify complementary stations from congestion levels at one station producing cross-station demand effects.

2 - Words Matter! Predicting Freelance Job Success Through Conversational Analytics

Jiannan Xu, Robert H. Smith School of Business, University of Maryland, College Park, MD, United States, Ashish Kabra, Kunpeng Zhang, Wedad Elmaghraby

As digital labor platforms become an increasingly important source of employment, understanding the factors that drive successful freelancer-client interactions is essential for promoting fair access to opportunities and improving labor market efficiency. This study applies conversational analytics and machine learning to analyze a unique dataset of tens of millions of messages exchanged between freelancers and clients. Drawing on Grice's Maxims (1975) from communication theory, it uncovers linguistic and behavioral patterns that influence freelance success. By developing AI-driven predictive models, this research advances natural language processing (NLP) applications in labor markets and provides actionable insights to help freelancers enhance their communication strategies.

3 - Real-Time Performance Feedback and Its Impact on Worker Performance

Zhi Chen, Zhejiang University, Hangzhou, China, People's Republic of, Yuqian Xu, Jayashankar Swaminathan, Weihua ZHOU, Yanbing Zhou

As digital technologies reshape the modern workplace, real-time performance feedback systems have gained traction as tools for enhancing productivity. While traditional feedback mechanisms provide periodic and infrequent updates, real-time systems leverage technology to deliver continuous, immediate performance updates throughout the workday. Although these systems hold promise and are increasingly adopted across industries, the rise of new technologies enabling real-time performance feedback raises critical and underexplored questions about how such continuous, data-driven systems influence employee behavior. Leveraging a unique dataset and quasi-experimental setting from Alibaba's Cainiao logistics warehouses, this study provides empirical evidence on the effects of real-time relative performance feedback systems. We find that real-time feedback led to a 7.0% improvement in efficiency without compromising quality. Notably, low-performing employees exhibited the largest gains, suggesting that frequent performance updates create stronger incentives for improvement among those with greater room for growth – potentially narrowing workplace performance gaps through upward productivity gains. At the same time, overall performance variability declined by 18.7%, though low performers experienced increased fluctuations. To explore the mechanisms behind these effects, we examine the system's dual role as both a feedback and monitoring tool. Importantly, we focus on the unique nature of feedback enabled by real-time systems, characterized by continuous updates on workers' rankings, individual efficiency, and comparisons to

warehouse averages. Positive feedback, such as improved ranking or exceeding the warehouse average, boosts effort and enhances performance, while negative feedback can discourage workers. Beyond feedback content, the awareness of being continuously observed fosters sustained improvements regardless of specific feedback signals.

4 - Eliminate Gender Wage Gap in the Gig Economy: a Solution from a Customer Service Platform

Xiao Ni, Fudan University, Shanghai, China, People's Republic of, Qiuping Yu, TIANJUN FENG

Job flexibility is a significant factor in gender wage gaps, yet wage disparities persist even in flexible gig work. While previous research has examined this issue in male-dominated, location-specific occupations like Uber driving, less is known about gender wage gaps in female-dominated or fully remote gig work. This study investigates gender wage differences among gig workers at an online chat center for a major retail platform, where pay is set by a transparent, gender-neutral algorithm. We find a gender pay gap in this female-dominated industry, with men earning 20% more per week and 2.4% more per hour. The hourly earning gap is primarily driven by men's faster work pace and higher chat concurrency, despite delivering lower service quality. However, once adjusted for experience, the gender pay gap disappears due to performance-based metrics (unbiased by gender) and the equalizing effect of remote work, which standardizes work timing across genders.

5 - Explain Worker Schedule Quality Disparities: Evidence from An American Restaurant Chain

Qiuping Yu, Georgetown University, Washington, DC, United States, Xiaoxuan Hou

While labor market disparity research has extensively explored racial and gender disparities in hiring, wages, benefits, and promotions, the disparities in the temporal quality of work schedules has largely been overlooked, despite its significant economic and social consequences for workers. Using large and granular shift data with roughly half a million unique shift records for over 2,400 unique customer-facing workers from an American casual restaurant chain, we examine gender and racial disparities in scheduling quality across two key dimensions: (1) sufficiency, which assesses the number of hours workers receive, and (2) predictability, which examines the frequency of last-minute shift changes. We address two main questions: (1) Are there gender and racial disparities in scheduling quality, and if so, what are they? (2) What mechanisms drive these disparities? Our findings reveal that women and minority workers are initially offered 3.3% to 10.7% fewer hours, and are 7.2% to 23.9% more likely to experience last-minute additions compared to White men, after accounting for human capital factors including role, tenure, intrinsic abilities, and absenteeism. We also show that classic explanations, such as workers' preference and availability, either do not hold true or only partly explain the observed disparities. By leveraging detailed field data, analyzing multiple dimensions of schedule quality, and systematically examining the mechanisms, we deepen the understanding of the drivers behind the gender disparities in schedule quality and underscore the importance of addressing these issues.

6 - Explain Worker Schedule Quality Disparities: Evidence from an American Restaurant Chain

While labor market disparity research has extensively explored racial and gender disparities in hiring, wages, benefits, and promotions, the disparities in the temporal quality of work schedules has largely been overlooked, despite its significant economic and social consequences for workers. Using large and granular shift data with roughly half a million unique shift records for over 2,400 unique customer-facing workers from an American casual restaurant chain, we examine gender and racial disparities in scheduling quality across two key dimensions: (1) sufficiency, which assesses the number of hours workers receive, and (2) predictability, which examines the frequency of last-minute shift changes. We address two main questions: (1) Are there gender and racial disparities in scheduling quality, and if so, what are they? (2) What mechanisms drive these disparities? Our findings reveal that women and minority workers are initially offered 3.3% to 10.7% fewer hours, and are 7.2% to 23.9% more likely to experience last-minute additions compared to White men, after accounting for human capital factors including role, tenure, intrinsic abilities, and absenteeism. We also show that classic explanations, such as workers' preference and availability, either do not hold true or only partly explain the observed disparities. By leveraging detailed field data, analyzing multiple dimensions of schedule quality, and systematically examining the mechanisms, we deepen the understanding of the drivers behind the gender disparities in schedule quality and underscore the importance of addressing these issues.

SA53

Building B Level 3 B313

Emerging Topics in Responsible and Sustainable Operations

Invited Session

MSOM: Sustainable Operations

Chair: Can Zhang, Duke University, Durham, United States

1 - Fair Online Hospital Diagnostic Service Scheduling: Helping Both Patients and Providers

Joyce Luo, Massachusetts Institute of Technology, Cambridge, MA, United States, Maureen Canellas, Dessislava Pachamanova, Georgia Perakis

The management of diagnostic services is a challenging and important task for hospital systems. Suboptimal scheduling of diagnostic services increases patient wait times and reduces the availability of overnight beds that could otherwise be utilized by incoming patients. Patients are served by a limited number of diagnostic technicians, who frequently experience uneven workload distributions and disproportionate workload demands that contribute to higher rates of exhaustion and burnout. We propose a framework for analyzing and addressing workload imbalances by developing algorithmic solutions to target contributing factors. In collaboration with a large academic medical center in the Northeastern US, we identify key factors related to total workload, load difficulty, perception of load fairness, and workload variability, and develop an online algorithm that optimizes daily diagnostic technician schedules to mitigate these specific factors. Our proposed algorithm balances technician workload considerations with patient and hospital needs, simultaneously incorporating considerations for patient urgency levels and characteristics. We provide analytical guarantees and analyze the algorithm's performance using echocardiogram order data from our hospital partner. Computational studies show that our online load-balancing algorithm effectively mitigates all identified workload imbalance contributors by improving load fairness for technicians by up to 28%, limiting difficult scan load, and reducing daily workload variability compared to current hospital policy – all without compromising operational performance in terms of patient throughput and wait time. Our algorithm's policy shortens patient wait time and improves throughput for certain patient types, especially for high demand scenarios.

2 - When Nonmateriality is Material: Impact of ESG Emphasis on Firm Value

Yakov Bart, Northeastern University, Boston, MA, United States, Sonam Singh, Ashwin Malshe, Serguei Netessine

This study examines the presence and the impact of emphasizing nonmaterial ESG factors on firm value. ESG factors are considered nonmaterial (or material) when excluding them from corporate disclosure would not (or would) significantly alter the overall information available to a reasonable investor. Drawing on agency theory, the authors argue that managers may divert their attention from firm value-maximizing activities by emphasizing nonmaterial ESG factors to enhance their reputations. The authors measure ESG emphasis by applying a deep learning model to earnings call transcripts of 5,079 firms from 2005 to 2021. They demonstrate that, on average, managers emphasize nonmaterial ESG issues more than material ESG issues. Furthermore, they find a positive association between nonmaterial ESG emphasis and media mentions of the firms. Notably, a 10% higher emphasis on nonmaterial ESG factors leads to a significant 10.7% decline in firm value. This negative association intensifies over time, is stronger for firms with higher free cash flow, and is weaker for firms with higher market share. Finally, the authors develop an interactive web app (https://bit.ly/esg_app) to assist executives and investors in comparing companies against competitors and quantifying the impact of shifting ESG emphasis on the firm value.

3 - The Value of Time- and Location-Commitment for Decentralized Emergency Medical Services

Pieter van den Berg, Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands, Andre Calmon, Andreas Gernert, Stef Lemmens, Maria Rabinovich, Gonzalo Romero

Emergency medical services (EMS) in many low- and middle-income countries utilize decentralized platforms coordinating independent ambulance providers. However, significant operational challenges arise from uncertainty in provider time availability and unpredictable idle locations. These uncertainties hinder reliable service coverage and negatively impact patient outcomes. Using data from our partner Flare in Nairobi, Kenya, we investigate the relative effectiveness of enhancing provider temporal

commitment versus spatial commitment to improve system coverage. We employ optimization models adapted for ambulance commitment uncertainty, a detailed case study analysis, data-driven simulations, and a game-theoretic model. Our findings quantify a stark "cost of decentralization": the coverage provided by Flare's 340 loosely committed ambulances could potentially be matched by fewer than 15 optimally deployed fully committed units. We find that enhancing spatial commitment generally yields higher marginal returns for improving coverage than solely increasing time availability. Adding just five optimized, location-flexible ambulances increased coverage substantially in simulation (e.g., by 5% over the baseline fleet) and reduced service variability. Simulations confirm the practical impact of interventions and validate model assumptions. For managers and decision-makers overseeing decentralized EMS platforms, prioritizing strategies that improve spatial coordination offers an efficient path to enhancing service reliability and performance. Actionable strategies include targeted incentives that encourage providers to relocate strategically or deploy a small fleet of location-flexible, platform-controlled units to fill critical coverage gaps. Our framework offers practical tools for managers to identify coverage gaps and assess the potential impact of such interventions in resource-constrained settings, ultimately aiming to enhance emergency response.

4 - Electric Vehicle Charging Station Access Sharing and Economics Implication

Yuan Ma, Ross School of Business, University of Michigan, Ann Arbor, MI, United States, Roman Kapuscinski, Ozge Sahin

This paper presents a comparative analysis of Tesla's Supercharger access strategies in the European Union (EU) and the United States (US), focusing on the expansion of access to non-Tesla electric vehicles (EVs). By developing a micro-foundation modeling framework, this research examines how these regional variations in charging infrastructure accessibility influence consumer utility, EV adoption rates, and competitive dynamics within the EV market.

5 - Mitigating Shortages of Generic Drugs: the Role of Reliability Certification

Xiaoyu Wu, Duke University, Durham, NC, United States, Robert Swinney, Can Zhang

Generic drugs play a vital role in the healthcare system. However, generic drugs have experienced severe shortages due to supply disruptions, caused by a number of factors, including low profit margins and a lack of information available to buyers about manufacturers' reliability, leading manufacturers to underinvest in supply reliability. Shortages result in numerous negative consequences for patients as well as institutional buyers of generic drugs, who may be forced to use more expensive alternative drugs when a shortage occurs. In this paper, we investigate the effectiveness of such a certification system. We propose a game theoretic model that captures the effects of generic drug manufacturer reliability certification on manufacturers, buyers, a GPO and patients. We find that certification may motivate manufacturers to improve their reliability; however, this may increase patients' shortage cost once the manufacturers adjust their wholesale prices and the buyers adjust their sourcing strategies in response. Nevertheless, we also identify conditions under which certification can benefit all stakeholders compared to the status quo with no certification. Finally, we find that combining certification with subsidies can further reduce the shortage cost when the subsidy level is sufficiently high, but it may backfire if the subsidy level is too low. Our findings offer insights for policymakers and nonprofit organizations in designing reliability certifications to mitigate generic drug shortages, highlighting when and why certification can benefit various stakeholders in the supply chain, while cautioning against unintended consequences arising from its impact on manufacturers' price competition and buyers' sourcing strategies

SA54

Building B Level 2 B216

Experimentation in Operations: Design, Estimation and Decision

Invited Session

Revenue Management

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

Co-Chair: Iavor Bojinov, Harvard Business School, Somerville, MA, United States

1 - Enhancing Efficiency and Robustness of Switchback Experiments: a Practical Model-Assisted Design

Tu Ni, Harvard Business School, Boston, MA, United States, Iavor Bojinov

Switchback experiments have become a popular alternative when conventional A/B testing is hampered by limits on user-level randomization or by interference between experimental units. In the presence of carryover effects, however, current switchback designs typically suffer from low statistical power and an unpredictable yield of observations, posing challenges in execution. This paper presents a novel model-assisted design framework for switchback experiments that guarantees a fixed number of valid observations, enabling more reliable and consistent experimentation. The proposed design introduces an adaptive mechanism that uniformly distributes observations across seasonal cycles, ensuring treatment-control balance and reducing sensitivity to peak periods. Additionally, the design incorporates refinements that adapt assignments more to past periods, creating temporal coupling between treatments and controls to enhance robustness, even in less predictable dynamic systems. We provide a comprehensive analysis of the design's bias and variance and develop tools to support statistical inference. Numerical studies indicate that our new design is markedly more efficient than existing approaches; in a realistic ride-sharing simulation environment, it reduces the mean-squared error by nearly 30%. This work contributes broadly to the field of experimental design under interference, bridging practical considerations with robust statistical methodologies.

2 - Optimal Targeting in Dynamic Systems

Yuchen Hu, Stanford University, Stanford, CA, United States, Shuangning Li, Stefan Wager

Modern treatment targeting methods often rely on estimating the conditional average treatment effect (CATE) using machine learning tools. While effective in identifying who benefits from treatment on the individual level, these approaches typically overlook system-level constraints that may arise when treatments induce strain on shared capacity. We study the problem of targeting in dynamic systems, where treatment decisions must be made one at a time as units arrive, and early decisions can impact later outcomes through delayed or limited access to resources. We show that optimal policies in such settings compare CATE values to delay-specific thresholds, where each threshold reflects the expected cumulative negative impact on the system of treating an additional individual at that delay level. We propose an algorithm that augments standard CATE estimation with off-policy evaluation techniques to approximate these thresholds from observational data. Theoretical results establish consistency and convergence guarantees, and empirical studies demonstrate that our method improves long-run outcomes considerably relative to individual-level CATE targeting rules.

3 - When Does Interference Matter? Decision-Making in Platform Experiments

Anushka Murthy, Stanford University, Stanford, CA, United States, Ramesh Johari, Hannah Li, Gabriel Weintraub

This paper investigates decision-making in A/B experiments for online platforms and marketplaces. In such settings, due to constraints on inventory, A/B experiments typically lead to biased estimators because of interference; this phenomenon has been well studied in recent literature. By contrast, there has been relatively little discussion of the impact of interference on decision-making. We focus on the commonly used frequentist hypothesis testing approach for making launch decisions based on data from customer-randomized experiments, and we study the impact of interference on (1) false positive probability and (2) statistical power. We obtain three main findings. First, we show that for monotone treatments -- i.e., those where the treatment changes booking probabilities in the same direction relative to control in all states -- the false positive probability of the naïve difference-in-means estimator with classical variance estimation is correctly controlled. This result stems from a novel analysis of A/A experiments with arbitrary dependence structures. Second, we demonstrate that for monotone treatments, when the state space is large, the statistical power of this naïve approach is higher than that of any similar pipeline using a debiased estimator. Taken together, these two findings suggest that platforms may be better off not debiasing when treatments are monotone. Finally, we numerically investigate false positive probability and statistical power when treatments are non-monotone, and we show that the performance of the naïve approach can be arbitrarily worse than a debiased approach in such cases.

4 - Experimenting on Markov Decision Processes with Local Treatments

Chonghuan Wang, Massachusetts Institute of Technology, Cambridge, MA, United States, Shuze Chen, David Simchi-Levi

Utilizing randomized experiments to evaluate the effect of short-term treatments on the short-term outcomes has been well understood and become the golden standard in industrial practice. However, as service systems become increasingly dynamical and personalized, much focus is shifting toward maximizing long-term cumulative outcomes, such as customer lifetime value, through lifetime exposure to interventions. To bridge this gap, we investigate the randomized experiments within dynamical systems modeled as Markov Decision Processes (MDPs). Our goal is to assess the impact of treatment and control policies on long-term cumulative rewards from relatively short-term observations. We first develop optimal inference techniques for assessing the effects of general treatment patterns. Furthermore, recognizing that

many real-world treatments tend to be fine-grained and localized for practical efficiency and operational convenience, we then propose methods to harness this localized structure by sharing information on the non-targeted states. Our new estimator effectively overcomes the variance lower bound for general treatments while matching the more stringent lower bound incorporating the local treatment structure. Furthermore, our estimator can optimally achieve a linear reduction with the number of test arms for a major part of the variance. Finally, we explore scenarios with perfect knowledge of the control arm and design estimators that further improve inference efficiency.

SA55

Building B Level 3 B314

Bridging Theory and Practice: New Insights in Optimizing Operational Decisions

Invited Session

Revenue Management

Chair: Mingliu Chen, The University of Texas at Dallas, Richardson, TX, United States

1 - Mergers and Product Repositioning: Theory and Empirical Evidence

Soo-Haeng Cho, Carnegie Mellon University, Pittsburgh, PA, United States, Zijun Shi, Xin Wang, Yushu Zeng

This paper investigates how mergers affect firms' decisions on product quality, variety, and pricing, as well as the subsequent implications for consumer welfare. We develop a game-theoretical model of a vertically differentiated oligopoly in which firms compete on both quality and price. Post-merger, the combined firm may consolidate its product lines to achieve cost synergies, potentially altering market outcomes. We show that the welfare impact of a merger critically depends on the quality tiers of the merging firms. Mergers between top- and mid-tier firms can enhance consumer surplus if quality improvements offset price increases. In contrast, mergers involving bottom-tier firms tend to reduce consumer welfare due to limited consumer valuation of quality upgrades and reduced variety. These findings challenge the conventional focus on price effects in merger analysis and highlight the importance of product repositioning. Empirically, we analyze mergers in the airline industry using a Multinomial Logit demand model and a coarsened exact matched difference-in-differences design. Our results validate the theory: mergers between top-tier carriers increase consumer surplus, while those involving bottom-tier airlines often diminish it.

2 - Price and Assortment Optimization Under the Multinomial Logit Model with Opaque Products

Jiaqi Shi, Columbia University, New York, NY, United States, Omar El Housni, Adam Elmachetoub, Harsh Sheth

An opaque product is a product for which only partial information is disclosed to the buyer at the time of purchase. Opaque products are common in sectors such as travel and online retail, where the car type or product color is hidden in the opaque product. Opaque products enable sellers to target customers who prefer a price discount in exchange for being flexible about the product they receive. In this paper, we integrate opaque products and traditional products together into the multinomial logit (MNL) choice model and study the associated price and assortment optimization problems. For the price optimization problem, we surprisingly show that uniform pricing is optimal which implies it has the same optimal pricing solution and value as the traditional MNL model. While adding an opaque product priced at the revenue-maximizing price may enhance revenue given arbitrary traditional product prices, this advantage disappears when all prices are optimized jointly. For the assortment optimization problem, we show that the revenue-maximizing assortment is nested-by-valuation for uniformly priced products. For non-uniformly priced cases, we propose a natural nested-by-revenue-and-valuation heuristic that performs extremely well in an extensive numerical study.

3 - Shared OR Solo? Platform Operations and Social Choices in Ride-Hailing

Jianfu Wang, College of Business, City University of Hong Kong, Kowloon, Hong Kong, Ming Hu, Zhoupeng Zhang, Hengda Wen

Ride-hailing platforms provide pooling services, allowing riders to share trips, which can alleviate driver shortages and reduce wait times—particularly during peak hours—but may compromise privacy, space, and security. We analyze a queueing model where riders decide whether to join the queue and, if so, choose between shared or solo rides based on past wait-time experiences. Shared riders may face additional delays while waiting for a carpooling match. We derive optimal pricing strategies to incentivize riders toward joining rates and sharing probabilities that maximize throughput, revenue, or social welfare. Our findings show that revenue-maximizing joining rates and sharing probabilities are consistently lower than those under social welfare or volume maximization. As the shared-ride experience improves (i.e., stochastic increases in sharing externality) while sharing likelihood remains stable (aligned with Chicago data), the revenue-optimal shared price rises, while the solo price surprisingly decreases—compensating solo riders for increased system congestion. Under social welfare maximization, the solo-shared price gap narrows compared to revenue maximization, yet prices are not necessarily lower. Finally, we validate our theoretical insights using ride-hailing data from Chicago in August 2019.

4 - Pricing Roses in Online Dating

Kasra Ganjizadeh Tari, University of Pittsburgh, Pittsburgh, PA, United States, Arian Aflaki, Michael Hamilton, Enfeng Xing

Dating platforms provide an opportunity to match with a partner. Many of these platforms monetize this interaction by offering virtual products that enhance a user's chances of matching. For instance, Hinge allows users to send virtual roses to others, and Tinder offers Super Likes. These virtual products can increase the likelihood of a successful match by placing the user's profile at the top of the recipient's match list and signaling strong interest in the candidate.

We study the problem of pricing these virtual products, which we refer to as “roses,” in a multi-period setting. Users decide whether to send a match request upon observing a candidate and whether to accompany their request with a rose. We consider three selling mechanisms: offering roses separately in each period, as a bundle, or through both options. In this context, we characterize user behavior and derive the revenue-maximizing pricing and design for offering roses. For each scenario, we examine platform revenue, customer surplus, and social welfare, and compare outcomes across these metrics.

5 - Ride-Hailing Platforms: A Near-Optimal Policy for Pricing, Matching, and Spatial Routing

Yang Bo, University of Toronto, Toronto, ON, Canada, Guangwen Kong

We study an infinite-horizon, stochastic, dynamic optimization problem for an on-demand ride-hailing platform. The decisions include a one-time service pricing decision at the beginning of the horizon, and the real-time driver-customer-matching and driver-routing decisions during the planning horizon. The objective is to minimize the long-run average profit across time. We first establish a theoretical upper bound on the average profit under any policy. We then design a near optimal policy whose theoretical performance gap with respect to this upper bound vanishes in a meaningful asymptotic regime. Numerical results corroborate the near-optimal performance of our policy.

SA56

Building B Level 3 B315

Market Design, Sustainability, and the Environment

Invited Session

Auctions and Market Design

Chair: Faidra Monachou, Yale University, New Haven, CT, United States

Co-Chair: Ilan Morgenstern, Yale University, New Haven, United States

1 - Offsetting Carbon with Lemons: Adverse Selection and Certification in the Voluntary Carbon Market

Ilan Morgenstern, UC Berkeley, Berkeley, CA, United States, Vahideh Manshadi, Faidra Monachou

TBD

2 - Electric Vehicle Mobility and Charging Behavior

Christian Kaps, Harvard Business School, Cambridge, MA, United States

The rapid adoption of electric vehicles (EVs) is transforming mobility patterns and creating new dynamics in energy systems. This paper investigates the interplay between EV user mobility behavior and the strategic value of flexible charging. Using empirical data, we analyze how driving routines, charging preferences, and spatial-temporal demand patterns influence both individual driver and utility-level outcomes.

3 - Delivery Terms for Voluntary Carbon Offsets

Safak Yucel, Georgetown University, Washington, DC, United States, Vishal Agrawal, Gokce Esenduran

In addition to abating their emissions, to reach voluntary decarbonization goals, corporations purchase carbon offsets under two delivery terms. The first is forward delivery, where the buyer orders a certain quantity before the seller undertakes the investment and yield uncertainty realizes. The second is prompt delivery, where the seller invests and uncertainty realizes, before the buyer orders offsets. Motivated by the importance of choosing the right delivery term for buyers and sellers, in this paper, we investigate economic and environmental implications of these delivery terms. Our results offer several managerial insights: First, a seller should prefer forward delivery if investment cost is low, but prompt delivery otherwise. Although one would expect that the delivery term that enables the seller to make a higher investment results in a higher profit, we find that prompt delivery may lead to a higher profit despite lower investment. Second, a buyer should prefer forward delivery only for a project with either a low or high investment cost. Interestingly, the delivery term that leads to a lower cost for the buyer may actually require higher abatement. Finally, environmental groups should promote forward delivery only for a project with a low investment cost. Moreover, when the buyer and seller prefer different delivery terms, the one preferred by the seller leads to greater emissions reduction, and should be promoted by environmental groups.

4 - Market Fragmentation and Inefficiencies in Maritime Shipping

Giacomo Mantegazza, USC Marshall, Los Angeles, CA, United States, Kostas Bimpikis, Salomon Wollenstein-Betech

Maritime transportation accounts for 90% of global trade, but ballasting—vessels traveling without cargo—imposes substantial economic and environmental costs. This paper examines the oil transportation industry, where approximately half of all miles traveled are sailed empty. While some ballasting is necessary due to inherent supply-demand imbalances in oil markets, our analysis demonstrates that market structure, specifically the fragmentation of vessel ownership, is also a primary driver, accounting for 10-20% of the total empty miles traveled depending on the market segment. In addition, we show that consolidating vessels into small shipping pools—sets of vessels operated under unified management—can reduce ballasting-related carbon emissions by up to 15%. This market-driven approach, which is gaining industry adoption, maintains competitive dynamics, given the limited scale of consolidation, while significantly improving efficiency. The gains arise from enhanced coordination within larger pools and expanded port coverage, reducing unnecessary vessel repositioning. More broadly, our findings quantitatively demonstrate that organizational changes alone—specifically, the consolidation of vessel operations—can generate significant environmental improvements by reducing empty miles. This provides a practical path toward sustainability that can complement and amplify the benefits of technological innovation.

5 - Impact of the Maintenance in Power Plants and Data-Driven Optimal Maintenance Scheduling

Abdullah Coskun, Stanford University, Menlo Park, CA, United States

This study examines efficiency degradation and maintenance optimization in U.S. coal power plants using comprehensive hourly production, efficiency, and maintenance data. Through a difference-in-differences analysis, we demonstrate that coal generators experience significant efficiency decline prior to maintenance, followed by approximately 1% efficiency improvement post-maintenance. We develop a theoretical maintenance model that establishes the optimality of constant-control policies under stationary parameters and provides adaptive strategies for non-stationary conditions through production cycle clustering. Our analysis reveals that corporate consolidation through mergers and acquisitions has led to a 30% improvement in maintenance scheduling accuracy. The findings advance our understanding of power plant operational optimization and maintenance effectiveness in the U.S. electricity sector, while contributing to the empirical literature on real-world efficiency metrics in maintenance policy evaluation.

SA57

Building B Level 3 B316

Social Media Analytics-Student Paper Contest

Award Session

Social Media Analytics

Chair: Lingshu Hu, Washington and Lee University, Lexington, VA, United States

1 - From Self to Other: Customer Reactions to Firm DEI Promotions on Social Media

Chen Jing, Boston University, Boston, MA, United States, Priyanga Gunarathne, Shuba Srinivasan, Abraham Seidmann

We examine how firms' DEI promotions on social media affect customer complaints. Analyzing millions of interactions from 15 U.S. firms on Twitter, we find that DEI promotions raise customer complaints and weaken the benefits of responsive service, unless aligned with customer identities. Our findings reveal tensions in digital DEI strategy and offer guidance for managing online customer relationships.

2 - Emotional Dynamics—the Nonlinear Impact of Sentiment Volatility on Review Helpfulness

Jiaqi Ren, Zhejiang University, Hangzhou, China, People's Republic of, Xixian Peng, Yifan Yu, Yutong Guo

This study introduces sentiment volatility as a novel affective cue in online reviews. Drawing on EASI theory and multi-method evidence, we reveal an inverted U-shaped effect on perceived review helpfulness, mediated by emotional contagion and conceptual processing fluency. Our findings offer insights for online platforms, reviewers, and merchants.

3 - Reuniting Forcibly Separated Families Through Shared Memories-

Huifeng Su, Yale School of Management, New Haven, CT, United States, Lesley Meng, Edieal Pinker

Social media is a critical channel for separated families seeking reunification but faces vast search spaces, incomplete data, and inaccurate memories. Using a large crowd-sourced social forum dedicated to family reunification, our proposed language model pipeline reduces search complexity and offers practical, scalable solutions for social media-based family matching.

4 - Optimizing Influence Campaigns: Nudging Under Bounded Confidence

Yen-Shao Chen, Yale University, New Haven, CT, United States, Tauhid Zaman

Online influence campaigns often fail because audiences resist messages outside their comfort zones due to bounded confidence. This research reveals how “nudging”—gradually shifting opinions—overcomes this resistance. Using control theory and Twitter data, we demonstrate how multi-agent systems can effectively polarize or unite audiences.

SA58

Building B Level 3 B317

Supply Chain Optimization

Contributed Session

Contributed

Chair: Poyraz Kayabas, Purdue University Fort Wayne, Fort Wayne, IN, United States

1 - Adaptation Strategies-Based Supply Chain Viability Optimization Under Long-Time Disruption Risks

Wei Pu, School of Business, Guangdong University of Foreign Studies, Guangzhou, Guangdong, China, People's Republic of

With the accelerated globalization of the appliance industry, an intertwined supply chain network has emerged through resource integration. However, the multi-tiered and interactive nature of this network significantly amplifies the risk of cascading failures across supply chain nodes, thereby exacerbating systemic vulnerability. To mitigate the adverse effects of long-term disruptions—such as economic losses and customer dissatisfaction—an efficient optimization model is required to enhance the viability of intertwined supply chains (SCV) by integrating agility, resilience, and sustainability.

This model aims to improve the viability of appliance supply chains through resource allocation optimization and adaptive strategy selection, focusing on the following key components:

- (1) Analyzing the dynamic leader (appliance firms)–follower (suppliers) and coopetitive decision-making structures in intertwined supply chains to identify viability measurement indicators from a complex network perspective;
- (2) Generating stochastic dynamic disruption scenarios using dynamic Bayesian networks and developing a multi-objective stochastic programming model;
- (3) Implementing strategies such as fourth-party logistics (4PL), backup facilities, safety stock, and self-healing mechanisms to enhance viability.

Using real-world data from China's small appliance industry during the COVID-19 pandemic, this study demonstrates the practical applicability of the proposed approach in improving SCV and addressing multi-objective stochastic programming challenges. The optimal results indicate an 11.23% reduction in costs and a 13.72% decrease in order fulfillment time. These findings contribute to the theoretical and methodological foundations of supply chain resilience optimization and offer theoretical insights and practical tools to enhance SCV under disruption risks in appliance supply chains, thereby promoting their sustainable development.

2 - A Framework for Retailers and Manufacturers to Maximize Their Profits Through the Utilization of Food Shelf-Life Extension Technology

Yue Yao, University of Tennessee, Knoxville, TN, United States, Mingzhou Jin

Our research addresses the critical gaps in understanding the economic trade-offs and operational adjustments required for the adoption of shelf-life extension technology. By analyzing the potential impact of shelf-life extension on production batch size, frequency, and associated costs, this research seeks to provide actionable insights for food manufacturers and retailers. Through modeling the relationship between manufacturers and retailers as a Stackelberg game, our research underscores the strategic interplay between these two entities. The manufacturer, acting as the leader, determines the extended shelf-life and wholesale price to maximize profit, while the retailer, as the follower, calculates the inventory cycle length and optimal order quantity to balance profitability against expiry losses. The analysis delves into the impacts of shelf-life extension on production batch size, frequency, and costs, providing actionable insights for stakeholders. The findings aim to inform policy and decision-making, demonstrating conditions where shelf-life extension technology can maximize the profits for manufacturer and retailer, thereby advancing sustainability in the food industry.

3 - Proposed Modification of Lithium Reverse Logistics Supply Chain Network: An Optimization Outlook

Oluwatosin Atitebi, University of Texas at Arlington, Arlington, TX, United States, Erick Jones

Increasing demand for renewable energy technologies to advance growing push for transition from carbon-intensive energy system to non-carbon-intensive energy system has brought to fore the overwhelming demand for lithium-ion batteries, one of the current viable energy storage alternatives. This alternative energy system however is material intensive, therefore, the continuous demand for lithium, consequently increasing mining activities for these materials while also creating an unintended environmental/health crisis due to increasing rate of end-of-life lithium-ion batteries (LIBs) and their accumulation. Lithium reverse logistics network is on an uptick trend in recent decade geared towards facilitation of lithium recycling from end-of-life and other sources of scrap LIBs to minimize cost, enhance recovery efficiency, and ensure resilience of the network. This paper is a concerted effort towards ensuring an optimized lithium reverse logistics network by understudying the comparative analysis between the traditional reverse logistics network and the proposed modified reverse logistics network using 2 deterministic optimization models of the reverse logistic network variants. Overall, the proposed modified reverse logistics network showed a tremendous cost reduction compared to the traditional reverse logistics network over the planning period of consideration, while carbon emissions of both networks is comparatively the same.

4 - Enhancing Semiconductor Supply Chain Resilience and Reliability: a Quantitative Modeling Approach Under Uncertainty

Poyraz Kayabas, Purdue University Fort Wayne, Fort Wayne, IN, United States

This study presents a quantitative modeling approach to address the resilience and reliability challenges of semiconductor supply chains. Due to their complexity, global geographical distribution, and vulnerability to disruptions, designing resilient and reliable semiconductor supply chains requires a strategic balance between cost efficiency and operational robustness. The proposed approach evaluates various disruption scenarios and analyzes system response behaviors to identify optimal supply chain configurations.

SA59

Building B Level 3 B318

Social Media and Behavioral Modeling for Public Health Decision-Making

Invited Session

Social Media Analytics

Chair: Hongru Du, Johns Hopkins University, Baltimore, MD, United States

1 - An LLM-Based Framework to Evaluate the Impact of Digital Media Exposure on Vaccination Choice

Shaochong Xu, Johns Hopkins University, Baltimore, MD, United States, Liyue Zhang, Amelia Jamison, Abe Hou, Hongru Du, Lauren Gardner

The digital information ecosystem plays a vital role in shaping public health behaviors, but the influence of social media–amplified exposure on individual health decision-making remains poorly understood. This study introduces a Large Language Model (LLM)–based framework to assess how health-related content consumed via social media informs vaccination attitudes and behaviors. Drawing on individual-level survey data and real-world news content from online platforms, our approach directly maps media preferences to vaccine attitudes, with applications demonstrated in both COVID-19 and MMR contexts. At the core of the framework is the hypothesis that individuals' attitudes and beliefs are reflected in the media content they consume. To test this, we create diverse "media diets" spanning the ideological spectrum, including politically extreme perspectives and verified misinformation, using actual circulating articles. The experiment proceeds in two steps: first, an LLM matches each participant to a personalized media diet based on their demographics and surveyed beliefs. Next, we evaluate whether this simulation allows the LLM to accurately predict vaccination choices as a function of individualized media exposure. Our findings reveal that LLMs using simulated media diets can closely replicate vaccine-related behaviors at the individual level, especially for those with well-articulated beliefs, underscoring a strong connection between personal beliefs and media consumption. These results significantly advance infodemiology by uncovering how information exposure mirrors underlying belief systems that shape health choices, offering valuable insights for designing targeted health communication. The generalizable framework provides a robust means to model debate and hesitancy around vaccination decisions beyond the current pandemic.

2 - Internet Search Trends and Infectious Disease Outbreaks & Understanding Spatial Variations In Relationship

Srikar Velichety, University of Memphis, Memphis, TN, United States, Utkarsh Shrivastava

Internet search trends (IST) are crucial for real-time surveillance of new cases of diseases (NCD). The predictive models based on IST are highly susceptible to the users' online search behavior, which bears similarities at the geographical level. The digital divide in underdeveloped nations has the potential to significantly skew the online search and disease outbreak correlation across geographical regions. This study investigates spatial variations in the association between IST and NCD, what factors contribute to this difference and how they can be incorporated to improve disease surveillance. We utilize a multilevel modeling approach to model the weekly data on

new cases of five infectious diseases in 48 states of the USA over five years. We consider the weather, state demographic, insurance coverage, tourism, and outbreaks in neighboring states. Our initial results suggest that internet search trends have a differential relationship to disease outbreaks across different geographies.

3 - Assessing Mmr Vaccination Strategies in Texas: A Scenario-Based Modeling Study of Measles Outbreaks

Kaiming Bi, UT Health Houston, Houston, TX, United States, Thuy Nguyen, Boya Peng

Measles, once common in the U.S., declined sharply after the MMR vaccine's adoption in 1971 and was declared eliminated in the U.S. in 2000. Measles has reemerged due to waning vaccination coverage and case importation and more than 200 cases reported in the past few months in West Texas. We used a calibrated mathematical model and stochastic simulations to assess the impact of vaccination on the expected number of cases. Raising vaccination coverage by 5% could prevent 133 cases (95% CI: 97–169), while a 5% reduction may result in 190 additional cases (95% CI: 133–273) in Gaines County. These findings highlight the critical need for robust vaccination strategies to mitigate measles outbreak risks.

4 - Identifying and Forecasting Importation and Asymptomatic Spreaders of Multi-Drug Resistant Organisms in Hospital Settings

Jiaming Cui, Virginia Tech, Blacksburg, VA, United States

Healthcare-associated infections (HAIs) from multi-drug resistant organisms (MDROs) pose a significant challenge for healthcare systems. Patients can arrive at hospitals already infected ("importation") or acquire infections during their stay ("nosocomial infection"). Many cases, often asymptomatic, complicate rapid identification due to testing limitations and delays. Although recent advancements in mathematical modeling and machine learning have aimed to identify at-risk patients, these methods face challenges: transmission models often overlook valuable electronic health record (EHR) data, while machine learning approaches typically lack mechanistic insights into underlying processes. To address these issues, we propose NeurABM, a novel framework that integrates neural networks and agent-based models (ABM) to leverage the strengths of both methods. NeurABM simultaneously learns a neural network for patient-level importation predictions and an ABM for infection identification. Our findings show that NeurABM significantly outperforms existing methods, marking a breakthrough in accurately identifying importation cases and forecasting future nosocomial infections in clinical practice.

5 - Evaluating Public Opinion on GLP-1 Drugs Using Social Media Analytics

Kristen Nixon

SA60

Building B Level 3 B319

AI in Healthcare and Medical Research

Contributed Session

Contributed

Chair: Khaled Salah, Khalifa University, Abu Dhabi, United Arab Emirates

1 - Scalable Metadata Curation via Context-Enriched LLMs and Semantic Term Guidance: A Case Study in Anatomical Standardization

Minoo Ahmadi, University of Southern California, Los Angeles, CA, United States, Minoo Ahmadi

Scientific data repositories face critical challenges in metadata standardization, particularly for domain-specific terminology that varies across research groups and experimental contexts. We present a novel automated framework that leverages large language models (LLMs) enhanced with contextual metadata to enable high-accuracy term curation. Each prompt incorporates information from multiple data fields within a record, providing rich semantic context that informs curation decisions and reflects the complex structure of scientific data.

We evaluate two guidance strategies that build on a shared foundation of contextual enrichment. The first, standard vocabulary prompting, appends the complete reference terminology to ensure comprehensive coverage. The second, similarity-guided prompting, uses semantic embeddings to retrieve the top-k most relevant vocabulary candidates. This similarity-based approach introduces a key innovation, offering substantial scalability advantages in repositories with large or expanding vocabularies. Building on these strategies, we develop a confidence-based hybrid method that dynamically selects the optimal prompting approach for each term based on semantic coherence assessment.

Evaluated on anatomical terms from the FaceBase craniofacial data repository, our proof-of-concept demonstrates that both approaches achieve high standalone accuracy, with the similarity-guided method exhibiting superior performance in capturing nuanced semantic relationships. The hybrid strategy further improves accuracy by intelligently combining both approaches. The framework effectively addresses terminology challenges, including synonyms, descriptive variants, and structural inconsistencies common in multi-source scientific repositories.

These results demonstrate that context-aware, LLM-guided semantic curation achieves robust accuracy in heterogeneous scientific databases, offering scalable solutions for metadata standardization with implications for data integration, interoperability, and research reproducibility.

2 - Synthetic Training for Assessment in Real-Time(Star): Enabling AI-Powered Patient Surveillance

Annajirao Challa, Binghamton University, Binghamton, NY, United States, Wen Cao, Sreenath Chalil Madathil, Mohammad Khasawneh

This research presents a novel AI-powered surveillance framework for monitoring oxygen therapy patients, with a dual focus on detecting mask dislodgement and early signs of perioral cyanosis. In clinical settings, visual cues often go unnoticed due to limited supervision. However, data scarcity, along with privacy, regulatory, and variability constraints, presents a major challenge to developing robust computer vision models. The Synthetic Training for Assessment in Real-Time (STAR) framework was designed to address these challenges. It introduces a hybrid synthetic data generation pipeline that combines procedural 3D modeling in Blender, GAN-based face generation via StyleGAN3 and CLIP, and digital illustration using Krita to simulate cyanotic symptoms. The visual fidelity was validated using Fréchet Inception Distance (FID), confirming their utility in downstream training. Semi-supervised annotation was employed to reduce manual labeling. Two state-of-the-art object detection models, YOLOv12 and YOLO-NAS were trained to detect oxygen masks. Eight lightweight CNNs were employed for cyanosis identification. Both detection and classification models were evaluated through baseline testing and k-fold cross-validation. YOLOv12 achieved a mean mAP@50 of 0.9859 ± 0.0019 while models like ShuffleNetV2 and ResNet18 exceeded F1-scores of 0.94. The results show that hybrid datasets, paired with these efficient neural architectures and robust evaluation, enable AI systems for patient monitoring. The STAR framework fuses generative and perception models, paving the way for ethical, scalable, and accessible AI solutions in data-sensitive environments.

3 - Llm-Augmented Medical Query Intention Detection with Interactive Bert

Shaochun Li, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Xuan Wei, Pengzhu Zhang

The proliferation of online medical platforms has fundamentally transformed patient consultation patterns, creating a critical need for accurate matching between patient queries and physician expertise. Each patient inquiry carries distinct medical requirements, making precise intention detection essential for optimal patient-provider pairing. This paper presents a novel framework that employs question answering (QA) data augmentation to enhance query intention detection in online medical systems. Our approach utilizes Large Language Models (LLMs) to generate semantically enriched answer data for patient inquiries, thereby augmenting the contextual information embedded within original questions. A modified BERT architecture, enhanced with an interactive attention mechanism, then integrates the generated answers with corresponding queries to capture nuanced semantic relationships and complex interactions. This methodology significantly improves medical query intention classification accuracy while maintaining adaptability to intention detection tasks across diverse online platforms. Comprehensive experiments on real-world medical platform data validate our method's effectiveness in enhancing both classification performance and service matching efficiency, providing a robust solution for understanding user intent in digital inquiry systems.

4 - Agentic AI Framework for Autonomous Clinical Trial Cohort Identification

Supraja Krovvidi, CapsicoHealth, Inc., San Mateo, CA, United States, Shloke Meresh, Siham Belgadi, Ramesh Jakka, Laurent Hasson

Clinical trial cohort identification remains a critical bottleneck in the pharmaceutical industry. Researchers must manually navigate complex medical coding systems, such as ICD-DX, ICD-PR, HCPCS/CPT, and DRG classifications, to define study populations. This manual approach to cohort definition is time-consuming, error-prone, and often lacks comprehensive coverage of relevant diagnostic and procedural codes. Recent developments in agentic AI have shown promise for drug development processes, but few implementations address the foundational challenge of accurate patient population definition. We hypothesize that an agentic AI framework integrating natural language processing, medical code validation, and automated cohort creation can significantly enhance the efficiency and accuracy of patient population definition for clinical research. We developed an agentic AI framework that integrates natural language processing, medical code validation, and automated cohort creation to enhance the efficiency and accuracy of patient population definition for clinical research. This work contributes to the expanding landscape of agentic AI in healthcare, addressing critical needs in integrating AI with clinical decision-making while incorporating safety and accountability considerations. Future research directions include expanding the framework to incorporate real-time clinical data integration, developing predictive models for cohort outcome assessment, and implementing advanced multi-agent architectures for autonomous study design optimization. Empirical evaluation demonstrates substantial improvements across key metrics. The framework achieves broader code coverage compared to manual methods, reducing cohort definition time by 95% while improving relevant patient capture by around 80%. Notably, for rare disease cohorts, the system attains 91% concordance with expert-curated definitions while discovering additional relevant codes overlooked by traditional approaches.

5 - Augmenting FDA Expert Committee Recommendations Using Large Language Models

Oliver Schaer, Drexel University, Philadelphia, PA, United States, Simon Spavound, Panos Markou, Nikolaos Kourentzes, Matthew Schneider

Institutions heavily rely on expert committees to advise on decisions requiring specialized knowledge by holding majority votes. However, by imposing binary decisions, nuanced information, such as uncertainty and emotion, is lost, which is integral to the experts' decision-making processes. For example, a member might vote in favor of a question but add a caveat or other important context that does not make it to the decision table where votes are presented. In this research, we utilize LLMs to capture advisory committee members' levels of uncertainty and emotion. Within the context of the FDA Advisory Committee, we specifically examine the decision to reject or approve the sale of drugs in the United States. We empirically test the incorporation of these nuances for improved decision-making to prevent costly product withdrawals after approval.

6 - Agentic Network and Knowledge Tree Based Medical Record and Multi-Modal Document Analyses and Comparison Across LLM Models for Pharmaceutical Sector in Areas Such as Patient Journey Analyses, Drug Development and Market Access

Jakka Sairamesh, CapsicoHealth, Inc, Palo Alto, CA, United States, Supraja Krovvidi, Siham Belgadi, Laurent Hasson, Shloke Meresh

Agentic-AI tools and Methods are fast becoming the foundation for automated decision making driven by the specific context-driven and task-based analyses and reasoning. In this presentation, we will focus on clinical medical record and document analyses using knowledge-networks and task-based reasoning and compare multiple LLMs (including LLAMA3, GPT4, Mistral, Gemini) on their performance to analyze and extract critical insights from records and documents. Such analyses has strong implications in the Pharmaceutical sector that deals with thousands of documents for drug-development, research, trials, applications for drug-approval, post-launch outcomes and more. We found that depending on domain and contextual searches (including text and image data) agentic methods combined with LLM accuracy can enable better research and improve clinical and pharmaceutical research and operations (e.g. MSLs, clinical affairs, clinical operations and more). We found over 85% accuracy when using agentic AI methods with reasoning when compared to RAG based context-specific searches. Our studies have shown that Agentic-AI based content-aware reasoning helps build automated systems of the future to drive better research and clinical operations, and improve healthcare outcomes.

7 - Enhancing Cross-Border Healthcare Data Exchange with LLMs, Genai, and Blockchain

Khaled Salah, Khalifa University, Abu Dhabi, United Arab Emirates

Healthcare systems across countries struggle to securely and efficiently exchange patient data due to fragmented regulations, lack of interoperability, inconsistent consent management, and limited trust among institutions. Existing solutions are unable to dynamically adapt to diverse legal, linguistic, and technical requirements, especially in cross-border scenarios. We propose a novel framework that uses LLMs, GenAI, and Blockchain technology to enable secure, regulation-aware sharing of healthcare data between institutions and across borders. LLMs and GenAI autonomously handle critical tasks such as interpreting local and international healthcare regulations (e.g., GDPR, HIPAA), verifying patient consent, selecting shareable data fields, harmonizing medical codes (e.g., ICD, SNOMED), and translating clinical content. Data is structured using FHIR standards, encrypted, and exchanged via a blockchain-governed access layer, which ensures traceability, policy enforcement, and tamper-proof auditability. The system addresses key challenges in global health data exchange: interoperability, compliance, privacy, and trust.

SA61

Building B Level 4 B401

Modeling and Uncertainty in Large Language Models

Invited Session

Quality, Statistics and Reliability

Chair: Raed Al Kontar, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Naichen Shi, University of Michigan, Ann Arbor, MI, United States

1 - Speak to the Policy: Refining Policies with Language Feedback

Qiyuan Chen, University of Michigan, Ann Arbor, MI, United States, Raed AlKontar

We propose a novel framework for refining decision-making policies using natural language feedback. In our approach, we have running statistical policy whose behavior can be continuously critiqued via language. This feedback is then encoded and grounded using a Large Language Model (LLM), which acts as an intermediary that interprets and distills the feedback into actionable policy updates. Our method bridges the gap between symbolic feedback and statistical learning, allowing policies to evolve through language-guided interaction.

2 - Inv-Entropy: a Fully Probabilistic Framework for Uncertainty Quantification In Language Models

Haoyi Song, University of Michigan, Ann Arbor, MI, United States

Large language models (LLMs) have transformed natural language processing, but their reliable deployment requires effective uncertainty quantification (UQ). Existing UQ methods are often heuristic and lack a fully probabilistic foundation, largely due to model limitations and knowledge gaps. This work begins by providing a theoretical justification for the role of perturbations in UQ for LLMs. We then introduce a dual random walk perspective, modeling input-output pairs as two Markov chains with transition probabilities defined by semantic similarity. Building on this, we propose a fully probabilistic framework based on a novel inverse model, which quantifies uncertainty by evaluating the diversity of the input space conditioned on a given output through systematic perturbations. Within this framework, we define a new uncertainty measure, Inv-Entropy, implemented using bootstrapping and Monte Carlo estimation. A key strength of our framework is its flexibility: It supports various definitions of uncertainty measures, embedding functions, perturbation strategies, and similarity metrics. We also propose GAAP, a perturbation algorithm based on genetic algorithms, which significantly enhances UQ performance in empirical studies. Extensive experiments show that Inv-Entropy outperforms existing semantic UQ methods. Finally, we introduce a new evaluation metric, TSU (Temperature Sensitivity of Uncertainty), which directly assesses raw uncertainty values, unlike traditional accuracy-based metrics that rely on correctness as a proxy.

3 - Trustworthy LLM-in-the-Loop Bayesian Optimization

Raed Al Kontar, University of Michigan, Ann Arbor, MI, United States

Bayesian optimization (BO) is a sequential decision-making tool widely used for optimizing expensive black-box functions. Recently, Large Language Models (LLMs) have shown remarkable adaptability in low-data regimes, making them promising tools for black-box optimization by leveraging contextual knowledge to propose high-quality query points. However, relying solely on LLMs as optimization agents introduces risks due to their lack of explicit surrogate modeling and calibrated uncertainty, as well as their inherently opaque internal mechanisms. This structural opacity makes it difficult to characterize or control the exploration–exploitation trade-off, ultimately undermining theoretical tractability and reliability. To address this, we propose LLMIBO: LLM-in-the-Loop BO, a hybrid framework for BO that combines LLMs with statistical surrogate experts (e.g., Gaussian Processes ($\mathcal{GP}$)). The core philosophy is to leverage contextual reasoning strengths of LLMs for early exploration, while relying on principled statistical models to guide efficient exploitation. Specifically, we introduce three mechanisms that enable this collaboration and establish their theoretical guarantees. We end the paper with a real-life proof-of-concept in the context of 3D printing.

4 - Iterative Reinforcement Learning for Inference Time Adaptation of Diffusion Models and LLMs

Rongbo Zhu, University of Michigan, Ann Arbor, MI, United States

Generative models, including diffusion models for visual synthesis and Large Language Models (LLMs) for text, show immense promise. However, efficiently guiding these models to produce outputs that adhere to complex criteria, such as underlying physical laws in video prediction, remains a significant challenge. This paper introduces Iterative Reinforcement Learning for Inference Time Adaptation of Diffusion Models (IRLIA), a versatile method using reinforcement learning to improve pre-trained generative models, with a primary focus on video diffusion and potential applications to LLMs, directly at inference time. IRLIA employs an iterative RL approach where a task-specific simulator (e.g., a physics experiment simulator) serves to construct the reward function. This guides the generation process via value functions and guided search, steering outputs towards desired characteristics like physical plausibility in videos. This method enhances model outputs without retraining the base model, offering an efficient performance boost. We plan to demonstrate IRLIA's effectiveness initially on physics-based video generation (e.g., fluid dynamics), showing clear improvements in physical plausibility and predictive accuracy. We will also discuss its potential applicability and preliminary investigations for guiding LLMs towards complex textual requirements.

SA62

Building B Level 4 B402

Adaptive Sampling and Monitoring in Spatially Distributed Systems

Invited Session

Quality, Statistics and Reliability

Chair: Jiahui Zhang, University of Wisconsin-Madison, Madison, WI, United States

Co-Chair: Kaibo Liu, UW-Madison, Madison, WI, United States

1 - Adaptive Sampling for Automated Post-Disaster Rapid Damage Assessment via Level-Set Cost-Aware Bayesian Optimization

Boyang Xu, Arizona State University, Tempe, AZ, United States, Hao Yan, Mostafa Reisi Gahrooei, Mohammad Ilbeigi

Natural disasters frequently inflict severe damage to the built environment, which demands a rapid, reliable, and cost-effective damage assessment for emergency response. However, traditional methods for Post-Disaster Damage Assessment often rely on static, labor-intensive data collection strategies that can be prohibitively expensive and struggle to adapt to dynamic post-disaster conditions. In this study, we propose a cost-aware Bayesian optimization framework that continuously guides autonomous data collectors, e.g., an Unmanned Aerial Vehicle (UAV), toward the most informative regions. By dynamically updating damage estimates across different geographic zones, our approach systematically reduces uncertainty while minimizing human resource needs and other operational costs. The result of the algorithm provides accurate and timely damage estimates that support informative and fast emergency responses.

2 - Active Learning for Data Stream Monitoring

Christian Capezza, University of Naples Federico II, Napoli, Italy, Antonio Lepore, Kamran Paynabar

In spatially and temporally distributed environments, supervised statistical process monitoring approaches require labeled data to identify in-control and out-of-control states. However, acquiring labels—such as through physical inspections or destructive testing—is often expensive and infeasible at scale. This work introduces a stream-based active learning framework for adaptive labeling under a constrained budget. The method integrates partially hidden Markov models (pHMMs) to exploit temporal dependencies and efficiently combine labeled and unlabeled data for multi-state classification. A novel dual-criterion strategy balances exploration, to identify previously unseen process states, with exploitation, to improve boundary resolution among known states. Unlike traditional pool-based approaches, our method operates in real time, issuing labeling decisions for each incoming observation. An online pHMM fitting algorithm is also developed with a tailored initialization strategy for imbalanced settings. The proposed active learning framework is validated through an extensive simulation study and a case study on monitoring a resistance spot welding process in automotive manufacturing. In this setting, dynamic resistance curves are monitored continuously, but labels are acquired through costly ultrasonic tests. Results show that the proposed active learning strategy substantially improves classification accuracy while minimizing labeling costs.

Acknowledgements: The research activity of C. Capezza and A. Lepore was supported by Piano Nazionale di Ripresa e Resilienza (PNRR) - Missione 5 Componente 2, Investimento 1.3-D.D. 1551.11-10-2022, PE00000004 within the Extended Partnership MICS (Made in Italy - Circular and Sustainable). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

3 - Active and Data-Efficient Spatiotemporal Monitoring via Adaptive Sampling Strategies

Jiahui Zhang, University of Wisconsin-Madison, Madison, WI, United States

Spatiotemporal data arise naturally in many physical and industrial systems, such as environmental sensing networks and distributed manufacturing processes, where sensor measurements are correlated both spatially and temporally. In real-world settings, sensor observations are often partial and sparse due to operational or resource constraints, making it challenging to monitor such systems in a timely and accurate manner. Existing approaches require either strong distributional assumptions or sufficient historical data, which limits their applicability in real-world applications where only sparse observations are available beforehand and the data distributions are unknown. In this work, we develop an online monitoring framework that integrates adaptive sampling with nonparametric modeling to efficiently and effectively detect abnormal behaviors under limited sampling budgets. A spatiotemporal Gaussian process is used to represent the system state, and is sequentially updated as new data arrive. An acquisition function inspired by Bayesian optimization principles guides sampling decisions at each time step, balancing uncertainty exploration and exploitation for potential anomalies. Simulation studies on synthetic spatiotemporal processes demonstrate the framework's ability to maintain high monitoring accuracy with constrained sensing resources and limited amount of historical data.

4 - Self-Expressive Generalized Linear Models for Federated Network Optimization

Mohammad Amini, University of Florida, Gainesville, FL, United States, Mostafa Reisi, Aleksandr Kazachkov

We propose a novel federated learning framework designed for decentralized networks with heterogeneous data, where each local client models its outcome variable using a generalized linear model (GLM). Unlike conventional federated learning, which assumes a global model shared across all nodes, our approach supports personalized model training while enabling localized cooperation. To do so, we leverage the topology of the underlying communication network to guide how models influence one another. This formulation allows our framework to interpolate smoothly between isolated local learning and centralized model fusion, depending on the network's structure and the degree of data heterogeneity. Moreover, the model includes self-tuning mechanisms that automatically adjust key federated learning parameters—such as regularization strength and neighbor influence—based on local and network-level signals. A key theoretical contribution is that we show standard federated learning is a special case of our framework when the network is fully connected. This positions our model as a strict generalization of classical FL, with added flexibility for personalization and structured collaboration. We develop a decentralized optimization algorithm based on the alternating direction method of multipliers (ADMM) to efficiently solve the problem without central coordination. Experimental results demonstrate that our method outperforms existing federated and personalized learning baselines in terms of predictive performance, adaptability, and robustness to network sparsity. Our framework is especially suited for applications such as precision agriculture, distributed healthcare, and sensor networks, where local models must adapt to their environment while benefiting from selective, topology-aware cooperation.

SA63

Building B Level 4 B403

Predictive and Prescriptive Analytics for Informing Decision-Making in Mission-Critical Systems

Invited Session

Quality, Statistics and Reliability

Chair: Bingjie Wang, University of South Florida, Tampa, FL, United States

Co-Chair: Mingyang Li, University of South Florida, Tampa, FL, United States

1 - A Multi-Stage Stochastic Optimization Approach for Maintenance Planning of Co-Located Infrastructures Under Deterioration and Budget Uncertainties

Bingjie Wang, University of South Florida, Tampa, FL, United States, Hung Nguyen, Qing Lu, Qiong Zhang, Mingyang Li

Deteriorating transportation infrastructure (TI) and water infrastructure (WI) are interdependent due to their co-location, with failures and degradation in one system influencing the performance of the other, while complex intra-system interactions further propagate deterioration impacts within each network. Given the inherent unpredictability of infrastructure deterioration and budget fluctuations, effectively managing these interdependent systems requires explicitly accounting for uncertainties. This research develops a network-wide, long-term joint maintenance planning framework for co-located TI and WI systems, capturing both intra-system and inter-system interdependencies. We further propose a multi-stage stochastic programming model to optimize maintenance planning, explicitly modeling uncertainties in deterioration rates of the co-located roads and pipes, as well as future budget variations over a defined planning horizon. A progressive hedging algorithm is employed to determine optimal maintenance decisions within this framework. A case study based on the Ybor City, Tampa testbed demonstrates the effectiveness of the proposed method, highlighting its capability for cost-efficient and proactive joint maintenance planning across interdependent infrastructure systems. Comparative analysis with alternative strategies further demonstrates the benefits of the proposed approach in reducing long-term costs and enhancing network reliability.

2 - Deep Neural Network for Predicting GPU Failure Time Distributions Under Competing Risks

Jie Min, University of South Florida, Tampa, FL, United States, Yueyao Wang

Time-to-event data, or lifetime data, are commonly seen in areas including biology, medical science, and engineering. In analyzing such data, one important topic is the prediction of the survival probability of units. This paper proposes a deep neural network for predicting the survival probability of units with censored data and competing risks without parametric assumptions. The neural network takes the sub-distribution hazard functions as outputs and uses likelihood loss with L0 smoothing penalty. The performance of the proposed neural network is evaluated and is compared with existing methods using simulated datasets. The advantages and disadvantages of the proposed and existing neural networks are discussed in terms of predictive discrimination and calibration. The proposed neural network is then applied to predict the survival probability of GPUs in the Titan supercomputer.

3 - An AI Enhanced Simulation Framework for Informing Nursing Home Staffing Decisions

Yulun Xu, University of South Florida, Tampa, FL, United States, Nazmus Sakib, Hongdao Meng, Nan Kong, Kelly Smith, Mingyang Li

Nursing homes (NHs) provide continuous care for frail older adults with complex needs, but rising demand due to population aging highlights the need for effective management strategies. Despite high costs, many NHs continue to face regulatory violations and workforce shortages, revealing systemic challenges in maintaining care quality.

Many of the existing works primarily assumed homogeneity among individual patients and analyzed service demand based on overall patient volume, overlooking the variation in care needs arising from diverse individual characteristics. To address these limitations and fill existing research gaps, we propose an AI-enhanced simulation framework for characterizing the heterogeneous service demand of NH residents over time and determining the appropriate staffing decisions. Specifically, deep learning methods will be integrated with digital simulation to predict service demand of a heterogeneous population of NH residents at both resident and facility levels, and further evaluate various types of staffing decisions, such as ratio-based and need-based staffing decisions. The effectiveness of the proposed work will be showcased via a real case study.

4 - A Comparative Analysis of Data-Driven Modeling Approaches to Forecast Cyanobacteria Algal Blooms in Eutrophic Lake Discharge Canals

Hung Nguyen, University of South Florida, Tampa, FL, United States, Mauricio Arias, Qiong Zhang, Osama Tarabih, Cassondra Armstrong, Detong Sun, Edward Philips

This research explores the use of data-driven models to develop forecasting tools for algal blooms (ABs) represented by the Chlorophyll-A (Chla) concentrations in the St. Lucie and Caloosahatchee canals in Lake Okeechobee, Florida. By different approaches, the study aims to understand the differences in Chla dynamics between the two canals, identify the main drivers of Chla concentration in each, and develop suitable Chla forecasting models for the canals' operation purposes. The findings not only contribute to improving understanding of water quality dynamics in Lake Okeechobee canals but also can provide flexible tools for the decision-makers in management strategies based on the waterbody's characteristics in general.

SA64

Building B Level 4 B404

Behavior in Queueing and Service Systems

Invited Session

Behavioral Operations Management

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

1 - Reference Dependent Preferences in Priority Services

Leon Valdes, University of Pittsburgh, Pittsburgh, PA, United States, Kaitlin Daniels

Service providers offer priority alternatives in a variety of settings, e.g., TSA lines at airports, faster pickup options in ride hailing, etc. In this work, we explore whether and how reference-dependent preferences affect customer choices in such settings, and as a result, how service providers should account for them when designing their menus of options. In particular, we consider the presence of multiple reference points---the quoted service times for the normal and fast service---and people's responses to them in priority settings, which are currently not well understood in the literature.

2 - Understanding Queue Abandonment: the Role of Wait-Time Communication and Distraction

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

Despite significant advancements in predictive analytics and service delivery enhancement, customers often still face long wait times in physical (e.g., retail, healthcare) and virtual environments (e.g., ridesharing, chatbots). Drawing on Maister's (1985) well-known principles, two predominant tactics have emerged: informing customers about their wait through clear communication, and employing distraction techniques to divert attention away from the wait. Any firm's attempt to find a balance by combining both tactics, but is this the best approach? We conduct a series of controlled experiments to examine how customer abandonment behavior is affected by how an organization informs or distracts its customers. We experimentally varied the ability to retrieve information and the frequency of information pushed amongst distractions.

3 - Silent Abandonment in Text-Based Contact Centers: Identifying, Quantifying, and Mitigating Its Operational Impacts

Antonio Castellanos, Hebrew University of Jerusalem, Jerusalem, Israel, Galit Yom-Tov, Yair Goldberg, Jae Park

To improve services, companies allow customers to interact with agents via text. Unlike traditional call centers, text-based contact centers face unique operational challenges due to information uncertainty, particularly in measuring customer experience proxies such as abandonment and patience. A major source of this uncertainty is *silent abandonment*, in which customers leave while waiting without notifying the system. As a result, the system is unaware that the customers left, and the agent's time is wasted. Furthermore, the system may not be aware (even retrospectively) whether a customer silently abandoned or was served.

We develop classification models to investigate the magnitude of silent abandonment. A sample of 17 companies shows that 3%-70% of the customers entering these contact centers abandon silently. A detailed text-based analysis of one company finds that 71.3% of abandoning customers do so silently, reducing agent efficiency by 3.2% and system capacity by 15.3%, incurring an estimated annual cost of \$5,457 per agent. Separate from this classification, we develop an expectation-maximization (EM) algorithm to accurately estimate customers' patience under information uncertainty and examine covariates influencing its value.

We find that companies should use classification models to estimate abandonment scope and our EM algorithm to estimate customer patience. We propose strategies to mitigate the operational impact of silent abandonment by predicting likely behavior or changing service design. Specifically, we show that while allowing customers to write while waiting in the queue creates missing data, it also significantly increases patience and reduces service time, leading to lower abandonment and staffing requirements.

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

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

As virtual waits replace physical queues, managers today must decide how to communicate wait characteristics to customers. We experimentally demonstrate that this decision can substantially influence customer behavior. Clock-framed waits, like Uber, convey information accurately but are unappealing. Conversely, people systematically underestimate the duration and disutility of queues.

SA65

Building B Level 4 B405

Predictive and Constraint-Based Learning Techniques with Business Applications

Invited Session

Artificial Intelligence

Chair: Neta Rabin, Tel-Aviv University, Tel-Aviv, Israel

Co-Chair: Gonen Singer, Bar-Ilan University, Ramat-Gan, Israel

1 - Resource Allocation in Classification Task with Feature-Based and Sample-Based Constraints

Gonen Singer, Bar-Ilan University, Ramat-Gan, Israel

Efficient resource allocation in classification tasks presents a significant challenge across various domains. These tasks demand high predictive accuracy and require compliance with complex constraints arising from resource limitations. Traditional machine learning (ML) models primarily focus on improving prediction accuracy but often fail to incorporate decision constraints, resulting in suboptimal resource allocation in real-world settings. This project proposes a two-phase framework that integrates classification with optimization models to allocate resources effectively while addressing two primary constraint types: (1) Feature-based constraints, which set restrictions on predicted classifications for a specific class based on certain feature values, and (2) Sample-Based Constraints, which require grouped samples to be assigned to the same class. To illustrate these constraints, consider a retention resource allocation problem for efficient churn risk management in a telecom company. The goal is to classify customers based on their churn risk level and design targeted interventions to reduce churn rates. Each customer receives a churn risk score based on their usage patterns, demographics, and interaction history. The operational challenge lies in efficiently allocating limited retention resources: there is a fixed budget for customer retention efforts across different customer segments (feature-based constraint), and certain groups (e.g., families on shared plans) must receive the same retention efforts to maintain fairness and prevent negative impact on other group members (sample-based constraint). Addressing this challenge requires balancing the prioritization of high-risk customers with the constraints of budget and operational limitations, ensuring the highest return on investment for retention efforts.

2 - A Graph-Based Sparse Filter Feature Selection Method for Multi-Class and Ordinal Time Series Classification

Neta Rabin, Tel-Aviv University, Tel-Aviv, Israel

Feature selection is a key pre-processing step for high-dimensional datasets. Filter-based methods evaluate feature relevance independently of classifiers. This work introduces a graph-based filter approach for multi-class classification that significantly reduces the number of features while preserving class-separability information. The method integrates the Jeffries-Matusita distance with diffusion maps for nonlinear dimensionality reduction. Features are ranked based on their distribution in the reduced space, allowing the selection of a compact subset with complementary separation capabilities. A modified version is also proposed to handle ordinal feature selection. Experiments on public tabular and time-series datasets show the method outperforms existing filter-based techniques.

3 - Detecting Coordination Within and Across Encrypted Platforms

Shahar Somin, Bar Ilan University, Ramat Gan, Israel, Tom Cohen, Jeremy Kepner, Alex Pentland, Keeley Erhardt

Adversarial activity on networks is escalating rapidly, with cybercrime damages projected to reach \$9.5 trillion annually in 2024. While machine learning and deep learning have been widely applied in cybersecurity, they often depend on large volumes of clean, labeled data and struggle to adapt to encrypted, fragmented, and evolving environments. Graph-based methods offer a structural alternative, but typically focus on explicit, observable links—limiting their effectiveness when connections are hidden or obfuscated. This reflects a fundamental divide between the data plane, where actions are recorded, and the control plane, where coordination actually occurs. To bridge this gap, I present a scalable framework for detecting coordination within and across encrypted platforms by leveraging latent temporal patterns. We introduce temporal fingerprinting, which characterizes user behavior through bursty activity dynamics. These behavioral signals enable both identity matching and community detection across domains, even in the absence of structural links or content similarity. We further propose a generative model that links control-plane coordination to emergent data-plane correlations, revealing the conditions under which structure-based techniques fail. The approach is adaptable to large-scale, content-restricted environments and supports applications in pre-emptive cyber defense, fraud detection, and multi-platform surveillance—providing new tools for AI-driven decision-making in adversarial and data-sparse settings.

4 - Data-Driven Surrogate Simulation Optimization for Firefighting Resource Allocation

Yaniv Leuchter, Tel Aviv University, Pardes Hanna-Karkur, Israel

Efficient planning and resource allocation are vital for firefighting departments to ensure rapid emergency response and operational readiness. As urban growth, climate change, and unpredictable crises intensify service demands, data-driven strategies are increasingly essential. This study addresses limitations in current decision-support tools, such as business intelligence systems, which often overlook dynamic interactions between fire stations and incidents. We propose a simulation-optimization framework to enhance resource allocation, aiming to reduce response times and strengthen system resilience. Our multi-phase approach begins with a high-fidelity simulation model built on multidimensional historical data—incident locations, response times, resource use, and station dynamics. This model supports evaluating alternative resource configurations. We then implement a simulation-based optimization algorithm to identify more effective deployment strategies. To accelerate computations, a surrogate model is developed using supervised learning, and dimensionality reduction techniques are applied to isolate key drivers of response efficiency. The framework also allows for sensitivity analysis and demand forecasting based on urban growth and climate trends. A case study in one district of the Israeli Fire and Rescue Service demonstrates the framework's effectiveness. Using historical data, we assess operational performance across multiple scenarios. Preliminary results show reduction in the average response time. This work informs strategic decisions for policymakers, including the Fire Commissioner of Israel, supporting evidence-based resource planning. Funded by the Ministry of Science and Technology, this research showcases the value of academic-government collaboration in improving emergency response systems.

5 - An End-to-End Imputing Autoencoder Framework for Artifact-Aware Sleep Apnea Detection

Rahul Sharma, Georgia Institute of Technology, Atlanta, GA, United States

Physiological signals, such as those used in sleep studies, are often affected by artifacts arising from patient movement, sensor displacement, or external interference. These artifacts can significantly degrade the performance of automated diagnostic systems, such as those for sleep apnea detection. Conventional approaches to mitigate this issue typically rely on manual artifact removal or supervised filtering techniques that require prior knowledge or expert-labeled data. However, these methods can be labor-intensive, non-scalable, and often fail to generalize across datasets with varying noise characteristics. In this work, we propose a novel unsupervised deep learning framework that addresses the dual challenge of artifact removal and apnea detection in physiological signals through a unified architecture. The proposed method utilizes a dual-purpose convolutional autoencoder designed to decompose multichannel physiological signals into clean and artifact components without requiring labeled artifact data. Alongside this decomposition, an integrated apnea detection module is trained directly on the clean signal representation to accurately identify apnea events. This joint training approach ensures that the apnea detection model learns to rely only on artifact-free features, enhancing its robustness and predictive reliability. The model is rigorously evaluated on both synthetic and real-world datasets. By simultaneously performing artifact removal and classification, our approach eliminates the need for separate preprocessing pipelines and enables end-to-end learning from raw physiological signals. This unified model offers a scalable and generalizable solution for artifact-resilient biomedical signal analysis, with promising implications for real-time and automated diagnosis in clinical and home-based monitoring settings.

SA66

Building B Level 4 B406

Pedagogical Design and Innovations for Analytics Education

Invited Session

INFORMS Committee on Teaching and Learning

Chair: Clara Novoa, Texas State University, San Marcos, TX, 78666, United States

1 - Teaching with the Burrito Optimization Game**Lawrence Snyder, Lehigh University, Bethlehem, PA, United States, Karmel Shehadeh, Luis Novoa**

Gurobi's Burrito Optimization Game is a fun, free, and interactive online game that aims to introduce learners to the basic ideas behind mathematical optimization (MO) in the operations research (OR), business analytics (BA), and supply chain management (SCM) fields. In the game, players drag and drop burrito trucks onto a city map in order to serve customers and maximize the profit of the burrito company. The underlying problem is a classic MO problem, a facility location problem. We conducted an experiment to explore the effectiveness of the game as an aid for teaching MO in a classroom environment. Our results show that the game is effective in both building understanding of and confidence in MO, as well as reinforcing players' prior knowledge of the concepts under proper instructor-guided discussion. In this presentation, we present an overview of the game, discuss its use in a classroom setting, and present our findings from this experiment.

2 - Purposeful Analytics Course Design**Julie Williams, University of West Florida, Pensacola, FL, United States**

The goals and benefits of prescriptive analytics assignments can be emphasized with innovative titles, scaffolded formats, and a template to guide students through the modeling process. In an online prescriptive analytics course, implementing these changes led to improved student engagement, with a higher percentage of participants submitting all modeling assignments—even when the lowest assignment grade could be dropped.

3 - Problem Solving and Panic Attacks: Teaching Analytics Without Frying Brains**Susan Palocsay, James Madison University, Harrisonburg, VA, United States, Luis Novoa, Scott Stevens**

In the never-ending quest to improve teaching and learning in our analytics classes, we are currently studying cognitive load theory (CLT). First introduced in the field of educational psychology almost four decades ago, CLT provides an architectural model of human cognition and processes for acquiring complex problem-solving skills. Based on this model, it identifies three types of cognitive load—intrinsic, extraneous, and germane—and gives guidance on ways to manage each type for optimal learning and retention of knowledge and skills. In this presentation, we will explain the core principles of CLT and explore their practical application in all three stages of teaching analytics, i.e., presentation, practice, and assessment. The goal is to foster *productive* cognitive effort, rather than “frying brains,” and enhance student performance outcomes while also maintaining mathematical rigor. We will consider how instructional methods that are informed by CLT have the potential to improve conceptual understanding and problem-solving ability in quantitative courses. Examples based on topics in introductory statistics and management science, where students typically struggle, will be used to illustrate these methods. We will also discuss ideas for pedagogical research to gather evidence in support of CLT in analytics classrooms.

4 - Build, Deliver, Showcase: Lessons from Scaling An Analytics Lab for Real-World Impact**Daniel Pirutinsky, UC Berkeley, Berkeley, CA, United States**

Teaching analytics effectively at scale demands innovative strategies for student engagement, real-world relevance, and pedagogical rigor. This talk outlines the evolution of the Master's of Analytics course, "Analytics Lab," a project-driven, capstone-style class emphasizing practical skills through the cycle of Build, Deliver, Showcase.

Launched in 2022, the course transitioned from traditional lectures to active-learning sessions, progressively refining student deliverables from basic reports to sophisticated dashboards and presentations. By integrating licensed industry datasets and structured external collaborations, the course enhances authenticity and professional relevance. A culminating public Showcase event—evaluated by alumni and industry judges—anchors student accountability and ensures high-quality outcomes.

This talk provides insights into structuring large-scale analytics education, fostering student professional development, and effectively leveraging real-world collaborations.

SA67

Building B Level 4 B407

Data-Driven vs. Rule-Based: Bridging the Gap for Real-World Applications

Invited Session

Data Mining

Chair: Jiajing Huang, Kennesaw State University, Atlanta, GA, 30346, United States

Co-Chair: Feng (Ryan) Lin, University of Washington, Seattle, WA, United States

1 - Deepmide: a Multivariate Spatio-Temporal Deep Learning Model for Ultra-Scale Offshore Wind Energy Forecasting**Feng Ye, Clemson University, Clemson, SC, United States, Xinxi Zhang, Michael Stein, Ahmed Aziz Ezzat**

The hyper-increase in scale and capacities of modern offshore wind turbines—with altitudes approaching many of the world's tallest landmarks—motivates a departure from univariate wind energy forecasting methods that are only focused on a single representative height. In response, we propose DeepMIDE—a multivariate statistical deep learning method which—unlike existing time-series and space-time methods—jointly models the spatial, temporal, and vertical variations and dependencies in the offshore wind resource and corresponding power output. DeepMIDE is a multi-input, multi-output integro-difference equation model with a nonstationary multivariate kernel, for which key physically meaningful parameters are learnt via an advanced deep transformer architecture that extracts the governing dynamics of wind propagation and formation from streams of weather maps. Tested on real-world data from the offshore wind energy areas in the U.S. East Coast—where Giga-watt scale offshore wind

farms are in-development— DeepMIDE is shown to produce forecasts that are of higher accuracy and enhanced interpretability than prevalent statistical and deep learning-based forecasting methods.

2 - Visual Analytics to Assist Human-Machine Teaming

Jieqiong Zhao, Augusta University, Augusta, GA, United States

Artificial intelligence and machine learning (AI/ML) models have become powerful tools for decision support, capable of processing vast amounts of data that would be otherwise impossible to utilize. Yet, in high-stakes domains, human experts are still relied upon to validate model outputs and make final decisions as no AI/ML model is perfect. Therefore, effective human-machine teaming requires humans to understand, trust, and appropriately leverage AI/ML predictions. In this talk, I present a human-centered approach to developing visual analytics systems that integrate expressive visualizations with interpretable AI/ML algorithms. I will share both theoretical and practical advancements aimed at calibrating human adoption of AI/ML models, including systems that support domain experts in real-world applications. I will also present findings from a controlled experiment that investigates how uncertainty visualizations influence trust calibration in human-machine interactions, offering insights into the design of more reliable and transparent decision support systems.

3 - Structure-Independent Prediction of Protein-Protein Binding Affinity via Sequence-Based Deep Learning

Lingtao Chen, Kennesaw State University, Marietta, GA, United States, Yixin Xie

Protein-protein interactions (PPIs) are fundamental for biological functions and play key roles in health and diseases, including cancer, neurodegenerative disorders, and infectious diseases. Binding affinity, measured as changes in Gibbs free energy (ΔG), is crucial for identifying PPIs. Accurate prediction of ΔG between protein pairs is therefore important for understanding biological mechanisms. Traditional structure-based approaches are often limited by the availability of high-quality experimentally determined structural data. Here, we introduce a fully sequence-based deep learning framework that integrates rich protein embeddings with multiple neural architectures to predict ΔG directly from primary protein sequences. We systematically evaluate convolutional, transformer, and graph-attention architectures in combination with custom and pretrained sequence embeddings. The best model variant demonstrates substantial gains in predictive accuracy and generalization over earlier methods. Our models not only deliver robust performance across a diverse benchmark of protein complexes, but also exhibit stability under sequence perturbations, which show their practical utility when structural information is unavailable. By bypassing the need for three-dimensional structures and achieving state-of-the-art performance from sequence alone, this framework paves the way for scalable, structure-independent exploration of PPIs in drug discovery, protein engineering, and beyond.

4 - A Comprehensive Database Collection for Alzheimer's Disease: Organizing Key Research Domains to Enhance Diagnostics and Treatment

Johaam Kathilankal Jis, Kennesaw State University, Kennesaw, GA, United States

Alzheimer's Disease (AD) is a neurodegenerative disease that is progressive and mainly affects the brain causing a decline in cognitive ability in the regions of memory, reasoning, and language. According to a study conducted by the National Institute of Aging, AD is currently the 7th leading cause of death in the United States. Our research goal is to review existing computational methods and databases for further AD research. The work first starts with a brief introduction about AD and then goes to discuss different computational methods within AD research and then goes on to discuss databases. The computational methods part of the work will focus on different methods that are useful in AD research which includes talking about different machine learning, natural language processing, and molecular dynamic simulations procedures. Then we go on to talk about databases which are classified into 8 different categories. Each category is then defined followed by an explanation of how each individual database within the category can help play a role in AD research. Finally, in the application section, we will talk about different applications related to AD treatment. This research ultimate goal is to aid medical professionals, researchers, and pharmaceutical companies with valuable data to improve AD treatment and research.

5 - Bridging the Gap: Costar-a for Smarter Prompting in Localized Language Models

Nzubechukwu Ohalet, Kennesaw State University, Atlanta, GA, United States

As data-driven approaches continue to advance the capabilities of artificial intelligence, optimizing prompt design has emerged as a critical factor in improving the performance of large language models (LLMs). In this work, we present COSTAR-A, an enhanced prompt engineering framework that builds upon the existing rule-based COSTAR method—which incorporates Context, Objective, Style, Tone, Audience, and Response—by introducing an additional component: Answer. While the original COSTAR framework effectively improves clarity and alignment for general-purpose LLMs, it demonstrates limitations when applied to localized or fine-tuned models. Through a series of controlled prompt-response experiments, we show that the data-driven COSTAR-A framework leads to more structured and decisive outputs in localized LLMs, such as Llama 3.1. These findings highlight how a flexible, data-informed approach to prompt design can bridge the performance gap between generic rule-based methods and the nuanced demands of real-world, low-resource AI deployments.

SA68

Building B Level 4 B408

Advancing Theories and Methodologies in Financial Data Mining: Quantile Models, Network Structures, Diagnostic Frameworks, and Social Signal Integration

Invited Session

Data Mining

Chair: Xiaorui Zhu, Towson University, ELLICOTT CITY, MD, 21252, United States

1 - Quantile Predictions for Equity Premium Using Penalized Quantile Regression with Consistent Variable Selection Across Multiple Quantiles

Shaobo Li, University of Kansas, Lawrence, KS, United States

This paper considers equity premium prediction, for which mean regression can be problematic due to heteroscedasticity and heavy-tails of the error. We show advantages of quantile predictions using a novel penalized quantile regression that offers a model for a full spectrum analysis on the equity premium distribution. To enhance model interpretability and address the well-known issue of crossing quantile predictions in quantile regression, we propose a model that enforces the selection of a common set of variables across all quantiles. Such a selection consistency is achieved by simultaneously estimating all quantiles with a group penalty that ensures sparsity pattern is the same for all quantiles. Consistency results are provided that allow the number of predictors to increase with the sample size. A Huberized quantile loss function and an augmented data approach are implemented for computational efficiency. Simulation studies show the effectiveness of the proposed approach. Empirical results show that the proposed method outperforms several benchmark methods. Moreover, we find some important predictors reverse their relationship to the excess return from lower to upper quantiles, potentially offering interesting insights to the domain experts. Our proposed method can be applied to other fields.

2 - Enhancing Bankruptcy Prediction: a Two-Layered Network Approach Using Latent Space Models

Tianhai Zu, University of Texas at San Antonio, San Antonio, TX, United States

In this study, we present a novel statistical approach to corporate bankruptcy prediction by leveraging complex network analysis. We introduce a two-layered network structure that captures both supply chain relationships and investment-co-investment patterns among companies, providing a more comprehensive view of corporate interdependencies than traditional methods. To analyze this complex structure, we develop a flexible multi-layered latent position model that efficiently extracts key features from the network. Our methodology employs advanced statistical techniques to estimate latent positions underlying this two-layered network, which are then utilized as predictors in a bankruptcy prediction model. Using the US public company data, we demonstrate that incorporating these network-derived features significantly enhances the predictive power of bankruptcy models. Our results reveal that these latent positions estimated from network structure capture crucial relational information that is highly relevant to a company's financial stability. This approach not only outperforms traditional prediction methods but also provides interpretable insights into the role of corporate interconnectedness in financial risk. Our work aims to offer a robust statistical framework for integrating complex relational data into predictive modeling for bankruptcy risk assessment.

3 - The Ripple Effect of Corporate Social Media: How Firm-Generated Content (FGC) Shapes Investor Sentiment from User-Generated Content (UGC) and Stock Performance

Xiaorui Zhu, Towson University, Ellicott City, MD, United States

The popularity of social media platforms has fundamentally transformed word-of-mouth communication, significantly impacting stock market dynamics. Using over 60 million posts from X.com (formerly known as Twitter), this study explores how the posts made by the company's official account (firm-generated content, FGC) versus those from general users (user-generated content, UGC) interact with firms' stock performance differently. With the computationally intensive framework that incorporates the panel vector autoregressive (PVAR) model, we find that UGC sentiment, particularly bullishness, has significant and positive influences on future stock returns, while FGC positive and negative sentiments do not exhibit strong direct effects on stock returns. However, our empirical findings confirm the significant indirect effects of FGC sentiments on the stock return mediated through UGC sentiment and engagement metrics. Specifically, a stronger positive tone in FGC significantly increases bullishness of UGC, which in turn raises stock returns. Likewise, negative sentiment from FGC does not directly influence stock returns but contributes to shaping bullish sentiment of UGC, which then negatively affects stock returns. The findings hold practical implications for both investors and firms, suggesting that strategic management of corporate social media communications can influence market perceptions and trading activities.

4 - A Unified Framework for Residual Diagnostics in Generalized Linear Models and Beyond

Zwei Lin, Texas State University, San Marcos, TX, United States

Model diagnostics is an indispensable component of regression analysis, yet it is not well addressed in standard textbooks on generalized linear models. The lack of exposition is attributed to the fact that when outcome data are discrete, classical methods (e.g., Pearson/deviance residual analysis and goodness-of-fit tests) have limited utility in model diagnostics and treatment. This paper establishes a novel framework for model diagnostics of discrete data regression. Unlike the literature defining a single-valued quantity as the residual, we propose to use a function as a vehicle to retain the residual information. In the presence of discreteness, we show that such a functional residual is appropriate for summarizing the residual randomness that cannot be captured by the structural part of the model. We establish its theoretical properties, which lead to the innovation of new diagnostic tools, including the functional-residual-vs-covariate plot and Function-to-Function (Fn-Fn) plot. Our numerical studies demonstrate that the use of these tools can reveal a variety of model misspecifications, such as not properly including a higher-order term, an explanatory variable, an interaction effect, a dispersion parameter, or a zero-inflation component. The functional residual yields, as a byproduct, Liu-Zhang's surrogate residual, mainly developed for cumulative link models for ordinal data (Liu and Zhang, 2018, JASA). In general, it considerably broadens the diagnostic scope, as it applies to virtually all parametric models for binary, ordinal, and count data, all within a unified diagnostic scheme.

SA69

Building B Level 4 B409

Empirical Research on Innovation and Technology

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Ran Zhuo, University of Michigan, Ann Arbor, United States

1 - How Big is too Big? the Competitive Effects of Shipping Technology Innovation

Anhua Chen, Cornerstone Research, Cambridge, MA, United States

Technological innovation drives economic progress but can also lead to unintended consequences, such as market over-consolidation. This paper examines the competitive effects of technological advancement in the container shipping industry, focusing on the relationship between vessel size and market structure, and its impact on welfare. Using a dynamic oligopoly model and proprietary data, we quantify the economies of scale associated with vessel size and explore how it interact with market competition and investment behavior. We find that a 10% increase in vessel size reduces operational costs by 3.4% but contributes to market concentration, potentially offsetting consumer benefits. Counterfactual analysis suggests that the welfare-optimal vessel size is around 20,000 TEU under current demand level, as larger vessels risk over-consolidating the market. Additionally, smaller innovation steps promote competition, while larger steps drive consolidation. These findings highlight the need for policymakers to balance the efficiency gains of technological advancements with their competitive impacts, particularly when designing infrastructure investments in industries with strong economies of scale.

2 - Human-Centric Perishable Inventory Management with AI-Assistance

Yu Nu, University of South Carolina, Chapel Hill, SC, United States, Meng Qi, Karan Girotra, Elena Belavina

Up to 20% of food purchased by restaurants is wasted before reaching customers, with commercial leftovers being the major contributor due to poor production planning. This study introduces two AI decision-making assistants that support perishable inventory management differently in human-centric commercial kitchens. We address the periodic review inventory control problem for perishable goods with a fixed shelf life and demand censoring in an offline, data-driven setting. One AI assistant leverages a data-driven prescriptive solution to the multi-period inventory control problem, directly telling how much a human decision maker should replenish for the upcoming season given past sales data. We justify this prescriptive co-pilot with associated performance guarantees. Building on the data-driven prescriptive solution, the second AI assistant is enhanced in terms of detecting potential human decision-making biases in managing perishable inventory. Using machine learning models, it identifies from past user behavior whether a human decision maker is biased in their inventory decision making, and (if so) what human bias likely accounts for. Through an online experiment with Prolific workers, we further investigate how human users react to the deployment of different forms of AI assistance and uncover factors influencing their effectiveness. Results show that both types of inventory co-pilots, whether providing data-driven prescriptive solutions or bias detection, improve perishable inventory management performance for human decision-makers. Additionally, integrating bias detection with prescriptive solutions could foster greater human adherence to algorithmic recommendations.

3 - The Effect of U.S.-China Decoupling on Investment in High-Tech Manufacturing

Ran Zhuo, University of Michigan, Ann Arbor, MI, United States

High-tech manufacturing is often characterized by rapid technology turnover, frequent and substantial investment, and significant economies of scale. Recent national policies emphasizing domestic self-reliance in the U.S. and China highlight potential interactions between these industry characteristics and national security considerations. In this paper, we study the effect of U.S.-China decoupling on investments in contract manufacturing capacity for semiconductor chips. Our unique dataset combines a comprehensive sample of worldwide plant-level capacity investments and a representative sample of global contract manufacturing orders for semiconductors on a quarterly basis from 2004 to 2015. We use these data to estimate key parameters in a structural model of contracting between manufacturers and their clients, as well as manufacturers' capacity investment decisions. Based on the estimated parameters and using counterfactuals, we explore a large global semiconductor manufacturer's potential responses to geographically specific national policies affecting investment incentives.

4 - Consumers Cart-Building Behavior in Online Grocery: a Structural Approach Using Generative AI

Sandeep Chitla, NYU Stern School of Business, Jersey city, NJ, United States, Srikanth Jagabathula, Vidyadhar Mudium

How do customers build their online grocery shopping carts across multiple product categories, and how do operational factors such as prices, delivery fees, and promotions influence their choices? Existing multi-category choice models in operations, marketing, and economics often overlook the dynamic and sequential nature of cart-building and struggle with scalability in real-world scenarios involving thousands of products. Machine learning models, while scalable, often function as black boxes with limited interpretability. Our approach bridges this gap by combining the scalability of machine learning with the interpretability of structural choice models. We model customers as agents who sequentially select product categories based on their needs and choose products within categories according to their preferences. Using Generative AI techniques, we estimate the utilities driving these choices, extending the application of generative AI beyond its conventional domains to build a customer digital twin that replicates behavior on online platforms. Our analysis leverages a dataset from a leading quick-commerce platform in India, comprising 8.5 million orders from 350,000 customers. Our model predicts the next product category added to the cart 6.6% of the time from approximately 2,000 categories, with the true category in the top-10 predictions 38.8% of the time. Structural insights reveal that customers respond more to delivery fee waivers than flat discounts of the same value, highlighting the behavioral tendency to avoid service fees. Counterfactual analyses show how increasing delivery thresholds impacts product demand, with discretionary items like chips and longer shelf-life items like onions seeing increased demand as thresholds rise.

5 - Dynamic Discrimination on Online Platforms

Peibo Zhang, Emory University, Atlanta, GA, United States, Xinyu Cao, Dennis Zhang, Ruomeng Cui

Research on discrimination has predominantly relied on audit studies, which provide clean causal estimates through randomized profile testing but offer only static snapshots of bias. We adopt a dynamic perspective to examine how racial discrimination evolves over time. Using observational panel data from an online educational platform, we find that African-American teachers receive 64.3% fewer daily bookings than White teachers during their first 12 weeks on the platform, despite offering a similar number of class openings. Employing individual fixed-effects models, we estimate how the booking gap between African-American and White teachers evolves over time. The results show a striking 954% widening of this gap over the 12-week period. Our mechanism analysis shows that new students are significantly less likely to book classes with African-American teachers than with White teachers, even when they have minimal or comparable reputation metrics. This initial disparity makes African-American teachers build their reputation metrics and customer base at a slower rate than White teachers. Over time, the growing gaps in reputation and customer base further exacerbate the booking gap, contributed by both new students and repeat students. Our findings suggest that traditional audit studies may significantly underestimate the extent of discrimination, and highlight how reputation and matching systems serve as discrimination amplifiers on online platforms.

Sunday, October 26, 9:30 AM - 10:45 AM

Building A Level 4 Auditorium

Welcome and Opening Plenary - Operations Research and Social Policy: Models that Can Make a Difference

Plenary Keynote

Plenary

2 - Plenary - Operations Research and Social Policy: Models that Can Make a Difference

Margaret Brandeau, Stanford University, Stanford, CA, United States

The United States faces numerous complex, intertwined social challenges such as rising income inequality, lack of access to healthcare, lack of affordable housing, homelessness, opioid addiction, and mass incarceration. How can OR inform good policies for addressing these and other social problems? Decision makers deciding which programs to invest in face a number of challenges, including limited resources, incomplete information about the potential effects of programs, and objectives that include not only welfare maximization but also economic, social, and political considerations. This talk describes modeling efforts in which OR has played and can play a role in informing such decision making. By providing a structured framework that uses the best available evidence, imperfect as it may be, and that captures relevant uncertainties, complexities, and interactions, OR-based models can be used to evaluate the potential impact of alternative public programs. We describe models focusing on opioid, housing, and criminal justice policies. We conclude with a discussion of useful lessons and opportunities for OR modelers who wish to work on problems related to social policy.

Sunday, October 26, 11:00 AM - 12:15 PM

SB01

Building A Level 3 A301

Daniel H. Wagner Prize

Award Session

Daniel H Wagner

Chair: Konstantina Mellou, Microsoft Research, Redmond, WA, United States

1 - Optimizing Multimodal Rideshare Pricing via Quasi-Convex Efficiency Objectives

Sajjad Taghiyeh, Lyft Inc., San Francisco, CA, United States, Hao Yi Ong, Zeyad Kassem

We introduce a novel optimization system developed at Lyft to automate strategic rideshare pricing. Replacing manual heuristics, the system uses a unique quasi-convex efficiency objective to align pricing across multiple product types with long-term financial goals like bookings uplift and ride growth. It generates granular price recommendations across all 300+ markets daily, balancing demand with supply and integrating causal inference models. This approach enhances pricing consistency and allows analysts to focus on strategy, yielding 1.7% incremental rides and significant efficiency gains.

2 - Rebalancing and Pricing of Near-Expiry Inventory in Online Grocery Retail

Long He, George Washington University, Washington, DC, United States, Ziyuan Zhou, Zhengling Qi, Guangrui Ma, Yuanbo Peng, Yushu Yao, Zhongyi Zha, Mingyong Zhao, Zhenyu Zhuo

We partnered with Meituan, China's largest on-demand service platform, to develop and deploy an integrated learning-and-optimization system for managing near-expiry inventory in online grocery retail. By coordinating nighttime transshipment and dynamic pricing across thousands of micro fulfillment centers, the solution reduced spoilage by 20% and improved profit contribution by 14%, delivering over 50 million CNY in net gains. This work demonstrates how advanced analytics can transform perishable inventory from a loss driver into a source of profit and sustainability.

SB02

Building A Level 3 A302

Emerging Topics in Sustainable Operations

Invited Session

MSOM: Sustainable Operations

Chair: Wayne Fu, University of Michigan-Dearborn, Dearborn, MI, 48126, United States

1 - On-TIME by Design - Ride-Pooling Subsidies for Workers

Wayne Fu, University of Michigan-Dearborn, Dearborn, MI, United States, WEE MENG YEO

In this study, we explore a situation in which a manufacturing company subsidizes a ride-pooling service due to unreliable or insufficient bus services. We develop a stylized analytical model and apply a game-theoretical approach to analyze the decisions made by workers, the ride-pooling company, and the manufacturer. The findings aim to provide insights for the ride-pooling service on when to enter the market and how to operate effectively, as well as guidance for the manufacturer on how to subsidize the service.

2 - The Impact of Present Bias on Social Welfare

Zhixin Liu, University of Michigan-Dearborn, Dearborn, MI, United States, Nicholas Hall, Yun Shi

This work challenges the common perception that present bias predominantly has negative effects on decision outcomes. By examining the effects of present bias on social welfare in a stylized economy, we explore various decision-making environments where the impact of present bias is beneficial. One such environment involves dynamic pricing decisions. Additionally, we demonstrate how present bias can refocus decision-making that is otherwise compromised by other behavioral biases, including sunk cost bias and value sensitivity bias. Finally, in a duopoly non-cooperative game, present bias can lead to an improved equilibrium solution.

3 - Transition to Green Products: Taxing Emissions Versus Taxing Products

Gal Raz, Ivey Business School at Western University, London, ON, Canada, Paolo Letizia, Wangcheng Yan

Governments worldwide are implementing environmental tax policies to incentivize green product adoption and reduce emissions. This study compares two established taxation strategies: one based on emissions generated by all products, whether green or “dirty” (emission-based taxation), and another specifically targeting the “dirty” conventional product (product-based taxation). Using a game-theoretical framework with Cournot competition, we analyze the decisions of two competing manufacturers serving consumers with heterogeneous product consumption preferences. Our findings reveal that while both policies can motivate green production when fixed costs are low, emission-based taxation may inadvertently incentivize manufacturers to revert to the “dirty” conventional product. In contrast, taxing the “dirty” product consistently encourages manufacturers to offer, and consumers to adopt, green alternatives. Further, when emission intensity is low, product-based taxation emerges as the optimal strategy for reducing emissions and curbing the production of “dirty” conventional products. However, at high levels of emission intensity, the effectiveness of either policy is contingent upon the tax rate and market equilibria. Finally, we propose a consumer-focused tax policy based on individual emissions, aligning with principles of environmental personal responsibility, and demonstrate its superior potential for emission reduction. These insights provide guidance for governments in designing effective tax policies tailored to product emission intensities and minimizing unintended consequences during the transition to green products.

4 - Environmental and Economic Impacts of Pooled vs. Separate Carbon Targets

Guowei Dou, Shenzhen University, Shenzhen, China, People's Republic of, Christian Blanco, Gokce Esenduran

Some firms set separate targets for their Scope 1 and 2 emissions, and Scope 3 emissions, while others set one target for entire emissions. We try to identify when firms prefer one target type over the other, and the implications of the firm's preference and decisions regarding overall emissions.

SB03

Building A Level 3 A303

Multi-objective Optimization: Theory, Algorithms and Applications

Invited Session

Multi Criteria Decision Making

Chair: Banu Lokman, University of Portsmouth, Portsmouth, United Kingdom

1 - Exploring Efficiency and Equity in Day-Ahead Electricity Auctions: a Multi-Objective Optimization Approach

Banu Lokman, University of Portsmouth, Portsmouth, United Kingdom, Gökhan Ceyhan, Murat Koksalan

European day-ahead electricity markets (DAMs) are two-sided auctions that determine next-day electricity prices by solving a market-clearing model to match supply and demand bids. These markets are conventionally designed to maximize total surplus across all participants. However, there is usually no market-clearing price that maximizes the surpluses of all participants simultaneously, leading to potential inequities. Specifically, some participants may incur negative surplus (losses) or forgo potential gains (opportunity costs). This study addresses these equity-efficiency concerns by developing a multi-objective optimization approach and derives insights that can inform more equitable market designs across all DAMs.

2 - Multi-Period Critical Resource Allocation via Mobile Facilities: a Multi-Objective Customer-Centric Approach

Ozlem Karsu, Bilkent University, Ankara, Turkey, Zehranaz Varol, Bahar Y. Kara, Francisco Saldanha-da-Gama

Allocation of critical and scarce resources is relevant in real-world applications such as disaster response operations, healthcare resource distribution, and long-term relief management. In such settings, fairness in service levels and accessibility to services or commodities are important criteria alongside efficiency-related ones. This problem is cast as a multi-criteria multi-period location-inventory problem with a coverage constraint. Three criteria are considered based on deprivation cost and distance or travel time. The model is then applied to an instance of an aid allocation problem following a disaster. The tradeoff between the different objectives is analyzed. The results show that the proposed models can better inform customer-centered decision-making processes in the face of limited critical resources.

3 - Local Upper Bounds for Polyhedral Cones

Firdevs Ulus, Bilkent University, Ankara, Turkey, Gabriele Eichfelder

The concept of local upper bounds plays an important role for numerical algorithms in nonconvex, integer, and mixed-integer multiobjective optimization with respect to the componentwise partial ordering, that is, where the ordering cone is the nonnegative orthant. In this study, we answer the question on whether and how this concept can be extended to arbitrary ordering cones. We define local upper bounds with respect to a closed pointed solid convex cone and study their properties. We show that for special polyhedral ordering cones the concept of local upper bounds can be as practical as it is for the nonnegative orthant.

4 - Convergence Rate Results for Stochastic Multi-Objective Linear Programs

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

Consider a stochastic multi-objective linear program in which the “cost” matrix is the expected value of a random matrix, the constraints are known and fixed, and we have access to a dataset or stochastic oracle capable of producing an unbiased estimator of the cost matrix. In this setting, we show that the vertices of the feasible set which belong to the efficient set are identified exponentially fast in the sample size. This result appears in contrast to the behavior of the Pareto set estimator, which converges at a slower rate according to a Central Limit Theorem.

5 - Scalarization and Approximation Methods for the Multi-Objective Set Covering Problem

Lakmali Weerasena, University of Tennessee Chattanooga, Chattanooga, TN, United States, Chathuri ththanayake

This study addresses the **Multi-Objective Set Covering Problem (MOSCP)**, an NP-hard challenge in multi-objective combinatorial optimization. While norm-based scalarizations have shown promise in approximating multi-objective minimization problems, ϵ -approximation algorithms for MOSCP remain underexplored. Our contributions are twofold:

- (1) We investigate the effect of using a common set of weight vectors across two scalarization methods—**weighted sum** and **weighted max-ordering**—under various conditions, employing four distinct weight generation strategies and solving both scalarizations exactly.
- (2) We extend theoretical insights by analyzing the **cost-effectiveness vector** from previous work and proposing a new version, leading to new approximation properties. Results show that the proposed weight generation methods perform well with an equal number of weights, and **weighted max-ordering scalarization** is recommended for generating robust initial solutions in two-phase algorithms.

6 - Risk-Averse Decision-Making Under Decision and Objective Space Uncertainty in Multi-Objective Combinatorial Optimization

Chathuri Aththanayake, Department of Mathematics, University of Tennessee at Chattanooga, Chattanooga, TN, United States, Lakmali Weerasena, Damitha Bandara

Uncertainty in input data significantly affects the quality of solutions in mathematical optimization, particularly when the computed solutions are not realized exactly. This study introduces a novel risk-averse decision-making approach when objective function coefficients are uncertain, which causes the decision space uncertainty in multi-objective combinatorial optimization problems. We construct a sensitivity region in the objective space based on reference solutions obtained from the deterministic formulation. To address the uncertainty in the decision space, we identify neighborhood solutions around the reference solutions that lie within this region, from which the risk-preference solutions are derived. We propose indices to quantify the risk associated with neighborhood solutions by capturing both decision space and objective space uncertainty. Additionally, we propose two decision supportive indices to assess the performance and risk levels of both reference solutions and neighborhood solutions. The effectiveness of the proposed approach is demonstrated in an instance of the Set Covering Problem, highlighting in enhancement of risk-averse decision-making under uncertainty.

Keywords: Multi-objective Combinatorial Optimization, Neighborhood, Risk-averse Decision-Making, Sensitivity Region, Uncertainty.

SB04

Building A Level 3 A304

Trading and Risk Management

Invited Session

Finance

Chair: Rita D'Ecclesia, Sapienza University, Rome, 00162, Italy

Co-Chair: Andrea Roncoroni, ESSEC Business School, Cergy-Pontoise, France

1 - Offsets vs. Calendar Spreads in U.S. Treasury Futures Markets

Michel Robe, University of Richmond, Richmond, VA, United States

Using regulatory data, we document that much of what is commonly labeled as “spreads” in U.S. Treasury futures (USTF) markets actually consists of non-canceling directional positions, known as “offsets.” For both leveraged funds and institutional asset managers, these offsets are substantial—roughly one tenth and one half, respectively, of their aggregate net exposures. Unlike calendar spreads, which are typically market-neutral, offsets carry directional risk and contract significantly during periods of market stress. This distinct behavior has important implications for prudential assessments and financial stability monitoring. To aid such analysis, we provide a table that allows market observers to approximate the gross exposures of four major USTF trader categories using public data.

2 - Integrating Energy-Related Risks In Bitcoin's Price

Rosella Castellano, University of Rome UnitelmaSapienza, Roma, Italy

Bitcoin's high return volatility and energy-intensive mining process raise growing concerns about energy costs and environmental impact. Since mining relies heavily on electricity, energy prices significantly affect both profitability and market dynamics. Understanding how energy fluctuations and investor sentiment interact with Bitcoin's fixed supply is key to modeling its price. We propose a framework integrating energy costs, sentiment, and protocol-driven supply into a market equilibrium model. On the demand side, investors allocate wealth across assets with sentiment and energy as shifters. On the supply side, miner behavior responds to price, energy, and network conditions. This model, supported by machine learning calibration, offers insights for policymakers and investors on how energy and sentiment shocks influence Bitcoin prices.

3 - The Effect of Accounting Restatements on Supply Chain Partner Acquisitions

Volodymyr Babich, Georgetown University, Washington, DC, United States, Gilles Hilary, Kai Wai Hui, Gouman She

Would accounting restatements, by revealing the true state of the firm, make the firm a more attractive acquisition target? Our empirical analysis shows that the answer is no for acquisitions in general, but yes for acquisitions by supply chain partners. Our model explains this finding by describing an equilibrium among firms in a supply chain, the investors, and auditors. The key feature is an asymmetry of information between supply chain insiders and outsiders that hides mismanagement at one of the firms in the supply chain. We explain how accounting restatements change investors' equilibrium beliefs and which forces may affect the likelihood of supply chain integration. The model allows features such as noise and bias in managers' reports, auditing technology investments, auditing efforts, and investors' bargaining power. Leveraging model predictions, we eliminate the alternative explanation that the market underpricing of the target firm is the major reason for such acquisitions. The difference in acquisition rates with versus without a restatement is greater when acquisition targets exhibit greater agency problems, when acquirers have better governance, and when acquirers cannot easily switch supply chain partners. Our paper guides managers in their relationships with supply chain partners and highlights the importance of accounting information in this context. In particular, managers should consider accounting information in their outsourcing and acquisition decisions.

4 - Fragility of Green Investments, An Analysis of the Risk Drivers

Rita D'Ecclesia, Sapienza University of Rome, Rome, Italy

This paper uses the FTSE Russell Green Revenues 2.0 model, aligned with the EU Taxonomy, to identify sustainable investments. We construct High and Zero Green Revenue (GR) portfolios within Climate Policy Relevant Sectors (CPRS) using data from 2013 to 2023 to examine the greenium—the premium for green assets—via a Green Minus Brown (GMB) factor. High GR portfolios outperform brown ones only during periods of market stress, with no consistent investor preference for sustainability. The greenium varies across sectors, underscoring the need for sector-specific analysis. We also compare GR-based portfolios with other ESG metrics, emphasizing the need for a consistent sustainability classification framework.

5 - Commodity Vendors: Operational Handling, Business Timing, and Integrated Risk Management

Andrea Roncoroni, ESSEC Business School, Cergy-Pontoise, France

Commodity businesses operate under exogenous market prices and product demand, and vendors seek to maximize performance by jointly optimizing order quantities and market timing. This study examines a stylized, risk-averse commodity vendor model and provides a comprehensive theoretical analysis of the decision-making problem, both in terms of standalone operational management and when combined with financial hedging (integrated risk management). Our analysis highlights the critical role of market timing as a valuable operational decision variable, which degenerates into a now-or-never outcome in the case of a conventional newsvendor. Furthermore, the study reveals an expansionary effect of integrated risk management on both quantity and timing decisions. Applications to the steel component industry motivate and illustrate the model.

SB05

Building A Level 3 A305

Starting Academic Career for New and Prospective Faculty in Data Science

Panel Session

Junior Faculty Interest Group

Co-Chair: Nathan Gaw, Air Force Institute of Technology, Wright Patterson AFB, OH, United States

Co-Chair: Hao Yan, Arizona State University, Tempe, AZ, United States

1 - Moderator Panelist

Nathan Gaw, Air Force Institute of Technology, Wright Patterson AFB, OH, United States

TBA

2 - Moderator Panelist

Nathan Gaw, Air Force Institute of Technology, Wright Patterson AFB, OH, United States

Launching a successful academic career requires navigating numerous challenges beyond research excellence, especially in the dynamic and rapidly evolving field of Data Science. This panel brings together experienced faculty members at various career stages to share insights and practical advice with new and prospective faculty (including senior PhD students and postdocs) aiming to build careers in this area. Panelists will discuss key aspects of the early academic journey, such as establishing a distinctive and sustainable research agenda in data science, securing initial funding, developing effective teaching strategies for data science curricula, mentoring graduate students, balancing interdisciplinary collaborations and service commitments, understanding the specific nuances of the tenure process for data science-focused faculty, and fostering a healthy work-life balance. The session aims to provide actionable strategies and address common concerns to help junior academics thrive.

3 - Moderator Panelist

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

4 - Panelist

Nick Street, University of Iowa, Iowa City, IA, United States

8 - Panelist

Kang Zhao, University of Iowa, Iowa City, IA, United States

9 - Panelist

Giulia Pedrielli, Arizona State University, Tempe, AZ, United States

10 - Panelist

Raed Al Kontar, University of Michigan, Ann Arbor, MI, United States

11 - Panelist

Yisha Xiang, University of Houston, Houston, TX, United States

SB06

Building A Level 3 A306

Doing Good with Good OR - Student Paper Competition

Award Session

Doing Good with Good OR

Chair: Vanitha Virudachalam, University of Illinois Urbana-Champaign, Champaign, IL, United States

Co-Chair: Shalome Hanisha Anand Tatapudi, University of South Florida, Tampa, FL, United States

1 - Co-Designing Solutions to Disrupt Sex Trafficking Recruitment with Lived Experts

Baris Tezcan, Northeastern University, Brookline, MA, United States

Using a transdisciplinary, community-based approach, we developed a Resource Allocation–Markov Chain model of how vulnerabilities dynamically influence sex trafficking recruitment. This provides determine how to most effectively invest in community programs to reduce recruitment. Our work has transformed counseling practices, influenced policy on survivor inclusion in research, and initiated the first longitudinal sex trafficking data collection—demonstrating how operations research can have meaningful social change when embedded in authentic community partnerships that value lived expertise.

2 - Learning Optimal and Fair Policies for Online Allocation of Scarce Societal Resources from Data

Bill Tang, University of Southern California, Los Angeles, CA, United States

We study the problem of allocating scarce societal resources to heterogeneous allocatees on a waitlist based on their covariates. Using historical data, our policy waitlists individuals for the resource maximizing the difference between their estimated outcomes and the resource dual-price. We show that our policy almost surely asymptotically achieves the expected outcome of the optimal out-of-sample policy. We evaluate our approach on allocating scarce housing in Los Angeles. Our policies can improve outcomes by 5.16% and that fairness notions can be achieved at a very low price of fairness.

3 - Swap: Empowering Community Collective Impact Through Innovative Resource Sharing

Weixiao Huang, Worcester Polytechnic Institute, Worcester, MA, United States, Andrew Trapp, Jennifer Pazour

The SWAP system enables nonprofits to efficiently share resources, addressing resource scarcity and fluctuating demands through multilateral-exchange optimization and virtual currency (re-)distribution algorithm. In practice, SWAP has facilitated over 65 exchanges, enhancing organizational effectiveness, and fostering community collaboration, which has pioneered a new model for nonprofit resource sharing. Nonprofits using SWAP report significant improvements in operational capacity and interorganizational trust, indicating its substantial potential for broader national adoption.

4 - Saving Lives: Data-Driven Optimization in Animal Shelter

Qiuxia Chen, The University of Texas at Dallas, Richardson, TX, United States

We worked with Dallas Animal Services to improve outcomes for shelter dogs. By analyzing patterns in how long dogs stay in the shelter and how different types of dogs are adopted, we developed a practical strategy to save more lives without requiring more space. Our approach helped increase the shelter's live release rate – defined as the proportion of dogs with positive outcomes - from 76% in 2023 to 86% in 2024. This framework is easy to follow, scalable, and can be adapted by animal shelters across the country.

SB07

Building A Level 3 A307

Resilience on Networks

Job Market Showcase

Job Market Showcase

Chair: Pelin Kesrit, Texas A&M University, College Station, TX, United States

1 - Deep Learning Model for Predicting Urban Flood Risk of a Public Transit System

Mubarak Iddrisu, University of Massachusetts Boston, Quincy, MA, United States

Flooding significantly disrupts public transit systems and disproportionately affects vulnerable transit-dependent groups such as older adults and low-income residents. Given increased transit vulnerabilities to flooding driven by climate change and rapid urbanization, accurate prediction models are essential for effective adaptation planning and equitable emergency responses. This study presents a deep learning framework integrating Geographic Information Systems (GIS) and convolutional neural networks (CNN) to predict flood susceptibility specifically for public transit networks in Boston. By leveraging spatial data encompassing topographic, hydrologic, climatic, and socio-demographic features, our CNN predicts flood impacts at high spatial resolutions. The model is trained using labeled data from floodplain maps and transit infrastructure exposure with 70% training and 30% validation data split. Further, we applied traditional machine learning approaches, including Support Vector Machine (SVM) and Random Forest (RF), for comparative analysis. Preliminary results indicate our CNN model achieves over 80% accuracy with an Area Under the Receiver Operating Characteristic (AUROC) exceeding 0.90, outperforming the traditional ML models. Our results identify transit routes and stops at high flood risk and facilitate targeted adaptation strategies. The study underscores the importance of incorporating socioeconomic vulnerability into flood prediction models to support equitable transit planning decisions. This integrated approach advances precise and equitable flood risk assessment and provides actionable insights for transit planners and policymakers to enhance resilience in flood-prone urban areas.

2 - Fortifying Distribution Network Nodes Subject to Disruptions

Pelin Kesrit, Texas A&M University, College Station, TX, United States, Bahar Cavdar, Joseph Geunes

We consider a distribution network for delivering a natural resource or physical good to a set of nodes, each of which serves a set of customers, in which flow disruptions may occur at one or more nodes. Each node receives flow through a path from a source node, implying that a node is disrupted if one or more nodes on each path from a source node experience a disruption.

All network nodes are vulnerable to a future disturbance due to a potential natural or man-made disaster, the severity of which follows some measurable probability distribution. For each node in the network, we wish to determine a fortification level that enables the node to withstand a disturbance up to a given severity level, while

minimizing the expected number of customers who experience a service disruption under a limited fortification budget. We formulate this problem as a continuous, nonlinear knapsack problem with precedence constraints, demonstrate that this optimization problem is NP-Hard for general tree networks and general disturbance severity distributions, and provide a polynomial-time solution algorithm for serial systems, which forms the basis for an effective heuristic approach to problems on tree networks. Our computational test results demonstrate the ability of the proposed heuristic methods to quickly find near-optimal solutions.

SB08

Building A Level 3 A310

The Economics of AI

Invited Session

Information Systems

Chair: Jisu Cao, Arizona State University, Gilbert, AZ, United States

1 - Rational Analysis of Curiosity and Information Modality in Video Representations

Jaewung Sim, University of Connecticut, Stamford, CT, United States, Eunsol Cho, Daegon Cho, Jiyong Eom

In the attention economy, capturing users' attention is crucial for promoting content on social media platforms, particularly for long-form videos where effective pre-consumption representation is vital. This challenge is complex due to numerous factors, including the interplay of multimodal information—how textual and image components in video representations influence users' decisions to consume content. In this paper, we address the question of how semantic congruence between image and text elements of video representations affects curiosity. To empirically examine this question, we collect and analyze video data from YouTube and elicit the semantic congruence between thumbnail images and titles using multimodal large language models. Our findings reveal that the impact of semantic congruence on video views varies based on the information format, such that the image-title similarity shows an inverted U-shaped relationship with viewership, while the text-title similarity presents a linearly decreasing impact. We conduct a randomized online experiment and find supporting evidence.

2 - Text-to-Image AI and the Novelty of User Generated Content on Social Media

Guangzhi Shang, Florida State University, Tallahassee, FL, United States, Jingchuan Pu, Yi Tong

Text-to-image AI tools enable social media users to transform their imaginative ideas into visual content in a wide range of visual styles. We argue that the AI tools also pose a spillover effect on the novelty of user-generated regular content on social media (i.e., content generated without an AI tool) because using AI tools could potentially change the reputation dynamics on social media platforms (i.e., reputation mechanism) and/or fundamentally reshape human cognitive processes and innovative thinking (i.e., AI-usage mechanism). Collaborating with an image-sharing social media platform, we employ a difference-in-differences (DID) approach to empirically demonstrate that using a text-to-image AI tool significantly reduces the novelty of user-generated regular images. That is, after using the text-to-image AI tool, AI users' regular images become more similar to the platform-wide recent images than those of non-AI users. We do not find evidence supporting the reputation mechanism. However, empirical evidence consistently validates the AI-usage mechanism.

3 - AI-Enabled Job Markets & Market Participation: AI Aversion and Jobseekers Expectations of How AI Shapes Competition

Sarah Bana, Chapman University, Orange, CA, United States, Kevin Boudreau

Artificial intelligence is increasingly mediating matching between employers and jobseekers, and potentially reshaping how labor markets function. In this paper, we investigate whether the use of AI alters jobseeker behavior—specifically, the willingness to participate in job markets. We argue that knowing AI is used to match candidates with jobs may change jobseekers' willingness to participate, insofar as it shifts their expectations about the job market competition they will face. To test this theory, we conducted a field experiment with 4,562 jobseekers, randomly assigning them to different disclosure conditions. When AI use was disclosed, jobseekers' participation was roughly 25% lower than both a human matchmaking treatment and a control group in which the source was not given. We find that participation responses varied systematically by the quality of match, and jobseekers' professional networks, in ways that are consistent with jobseekers forming expectations about how AI affects the scale of competition, match prediction quality, and the information and sources used to generate predicted matches. Thus, the “AI aversion” in this context appears to be related to how AI shifts expected payoffs from participating. We conclude with implications for platform design, policy, and the future of AI-mediated labor markets.

4 - Can Kindness be Matched? Dynamic Match and Goal Match in Civic Crowdfunding

Xue Tan, Southern Methodist University, Dallas, TX, United States, Haoyan Sun, Oliver Yao

Match funds, funds supplemented by private or public institutions (“matchers”) to match charitable contributions made by individual donors, are widely adopted in traditional fundraising practices. Two match fund schemes prevail: a dynamic match fund is a commitment to match each contribution at a given ratio, and a goal match is a commitment to supplement a pre-announced amount of funds upon goal attainment. Matchers have a natural tendency to choose dynamic match over goal match, especially on crowdfunding platforms, because dynamic match funds enable higher visibility – upon every individual donation, a dynamic match fund donation will be transacted and displayed for public consumption, while a goal match fund donation is usually a one-time lump-sum contribution. By leveraging over ten years of data from a civic crowdfunding platform that supports the United Nations' sustainable development goals (SDGs), we uncover the differential impacts of dynamic match and goal match on fundraising performance. While the added match funds could always guarantee a higher total pledge, only dynamic match but not goal match crowds out total individual contributions. A more granular analysis further revealed that dynamic match funds reduce both donor participation and average contribution size. Surprisingly, goal-match funds led to an increase in average contribution size, offsetting its negative impact on donor participation. These findings contribute both to the theoretical understanding of match funds and add insights to the fundraising literature in economics and the crowdfunding literature in information systems.

5 - Balancing Safety, Engagement, and Compliance: Evaluating the Impact of Nsfw Governance on User Behaviors for Gen-AI Platforms

Jisu Cao, Arizona State University, Tempe, AZ, United States, Keran Zhao, Xiaowei Liu, Che-Wei Liu, Jiang Duan

Generative AI (Gen-AI) technology has rapidly transformed the content creation industry, enabling users to produce text, images, and videos with unprecedented efficiency and accessibility. While the new technology unleashes human productivity and creativity, it also presents the risk of misuse, particularly in generating NSFW content. Strengthening NSFW content governance creates a tradeoff between fostering a safer digital environment and maintaining user engagement. On one hand, strict governance protects users from harmful, exploitative, or distressing material, ensuring legal compliance and preserving platform integrity. However, such governance may lead to reduced user autonomy and increased dissatisfaction among certain user groups who value creative freedom. By leveraging a natural experimental policy change, where a Gen-AI platform tightened its NSFW policy in March 2024, we investigate how users with prior NSFW content intentions responded to the increased moderation. We find that introducing a strict NSFW filter led to a 0.7% decline in the subscription rate for users with NSFW intention, indicating a potential economic cost for platforms. Our findings highlight the complexities of NSFW content governance and offer important managerial implications for Gen-AI platforms.

SB09

Building A Level 3 A311

Preserving Privacy

Job Market Showcase

Job Market Showcase

Chair: Yuxin Liu, University of Pittsburgh, Pittsburgh, 15217, United States

1 - Joint Differential Privacy for Resource Allocation: Protection Capability, Price of Privacy and Algorithm Design

Du Chen, Nanyang Technological University, Singapore, Singapore, Geoffrey Chua, Victor Jose

Resource allocation is a fundamental problem in management science. Recent growth in privacy concerns has made this problem more challenging because private data might be inadvertently revealed through the optimal allocation decisions. While differential privacy (DP) limits data leakage in resource allocation, it often compromises decision utility substantially. To address this, we adopt joint differential privacy (JDP), another data privacy standard and provide a new comprehensive understanding of this approach. We formally show that JDP provides privacy protection as strong as DP if recipients do not disclose their allocations. Moreover, JDP achieves this with substantially lower utility loss, with per-recipient loss vanishing as the system scales. However, the direct application of existing JDP algorithms may not reap the full benefits of lower utility loss and could empirically underperform compared to DP. To address this, we exploit the primal-dual relationship and develop the Noisy Dual Mirror Descent (NDMD) algorithm, a minimax near-optimal approach for a large class of convex resource allocation problems. NDMD offers a practical solution for privacy-preserving allocation problems, including the long-debated public policy problem of allocating Title I US education grants with privacy. Our simulations using real-world data show that shifting from DP to JDP reduces nationwide misallocation by as much as 52% and significantly mitigates the disproportionate impact of privacy protection on historically disadvantaged and underrepresented minority groups. Our results highlight the importance of operational knowledge such as disclosure discretion, downstream operational tasks, and algorithm design in balancing privacy with utility.

2 - Privacy-Preserving Data-Driven Inventory Management

Lorraine Yuan, Cornell Tech/Cornell University, New York, NY, United States, Elena Belavina

Firms have access to increasing amounts of customer data, which facilitates better inventory management. However, security researchers have demonstrated that attackers can recover sensitive data by combining data-driven decisions with background information. This study develops and compares privacy-preserving approaches for data-driven inventory management that are robust to such attacks. We propose differential privacy-based adaptations of two leading data-driven methods—the two-step estimate-then-optimize and the one-step joint estimate-optimize paradigm—and establish theoretical guarantees on their performance. Our analysis characterizes a firm's tradeoff between privacy protection and profitability. It compares the three algorithms in terms of the difficulty of privacy preservation, profit losses from privacy, the optimal privacy protection levels, and customer surplus. We show that the one-step paradigm dominates the two-step paradigm by enabling lower-cost privatization through more targeted noise injection, achieving lower privacy exposure risk, higher profitability, and greater customer surplus. This reshapes our understanding of privacy protection—the costs and benefits of privacy preservation are not only driven by the privacy algorithms used but are also fundamentally shaped by the underlying data-driven decision making paradigm. This implies that privacy-conscious firms should deploy the private version of the one-step data-driven algorithm. We show that a firm that uses our methods and chooses an optimal privacy level will increase its demand, profit, and customer surplus (under most conditions)—challenging the notion that privacy protection necessarily comes at a cost. Our numerical analysis reveals that one-step algorithms offer 3 to 6 times stronger privacy protection and around 30% higher profits than two-step algorithms.

3 - Differentially Private Sequential Learning

Yuxin Liu, University of Pittsburgh, Pittsburgh, PA, United States, Amin Rahimian

In differentially private sequential learning, agents inject endogenous noise into their actions to preserve privacy. This leads to markedly different outcomes depending on whether signals are continuous or binary. For continuous signals under a finite privacy budget $\epsilon \in (0, 1)$, we propose a novel smoothed randomized response mechanism that adjusts the noise level based on a signal's proximity to a decision threshold. Unlike traditional randomized response, which applies uniform noise, our approach allows actions to more accurately reflect both private signals and public history. This enables faster information aggregation, achieving an accelerated learning rate of $\Theta(\epsilon \log(n))$, compared to the slower $\Theta(\sqrt{\log(n)})$ rate in the non-private regime where $\epsilon \rightarrow \infty$. Moreover, in the non-private setting, the expected stopping time for the first correct decision and the number of incorrect actions both diverge, while under privacy constraints, these metrics remain finite. This reveals a surprising efficiency gain in private continuous-signal settings, where the variable flipping probabilities amplify the log-likelihood ratio across time. In contrast, binary signals suffer under privacy: the use of constant randomized response before an information cascade reduces the informativeness of actions, significantly hampering learning. These findings highlight a counterintuitive insight—privacy can enhance learning with continuous signals but impairs it with binary ones.

4 - Privacy with Information Externalities and Complexity

Qingyi Wang, School of Information, University of Michigan, Ann Arbor, MI, United States, Eytan Adar, Yan Chen

Privacy decisions in social communication are complex in that they affect the individual making them and often everyone that person is connected to. Investigating information sharing with externalities in an abstract information-sharing game in the lab, in real life involving friends, and on a social media platform, we find that participants overshare relative to the Bayesian Nash equilibrium prediction in the game, and that externalities increase the likelihood of sharing in both the game and real-life sharing involving friends. When we exogenously reduce computational complexity in the game by providing an average payoff matrix, participants are 19.4 percentage points more likely to best respond, leading to a 24.4% increase in efficiency. Furthermore, more risk-averse individuals share less and therefore perform better in the game. The tendency to share sensitive information involving friends is negatively correlated with both risk aversion and ambiguity aversion. Our results highlight the potential to improve users' welfare by reducing the complexity in their privacy decisions.

SB11

Building A Level 3 A313

Disease Screening and Treatments

Job Market Showcase

Job Market Showcase

Chair: Sayaka Ishizawa, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, United States

1 - Universal Lung Cancer Screening Guidelines Under Heterogeneous Patient Responses

Raul Garcia, Rice University, Houston, TX, United States, Andrew Schaefer, Iakovos Tournazis

Health organizations (*society*) generally prefer to recommend universal screening guidelines to at-risk individuals (*patients*) for various diseases, for which coverage is typically provided by insurers. However, patients are heterogeneous both in disease risk and disutility associated with screening, resulting in varying health outcomes. Infrequent recommendations leave at-risk patients to decide if to pay out-of-pocket in the interim, while frequent recommendations may result in unnecessary screening disutility for low-risk patients and unnecessary costs to society. Moreover, patient adherence to guidelines remains low for many diseases, further complicating the health benefits and costs realized in practice. We address these challenges in the context of lung cancer screening for eligible ever-smokers and employ a bilevel optimization model to compute universal lung cancer screening guidelines that maximize the societal net monetary benefit (NMB). With *society (leader)* determining the covered universal screening schedule, representative patient classes (*followers*) are modeled as partially observable Markov decision processes (POMDPs), maximizing their own NMB with respect to their smoking level, screening disutility, willingness-to-pay for out-of-pocket costs, and adherence rate. We provide analytical results for the POMDP and conduct numerical experiments.

2 - Optimizing Vascular Access Timing in Pre-Dialysis Patients: a Markov Decision Process Approach

Yiwen Cao, University of Southern California, Los Angeles, CA, United States, Sze-chuan Suen

Effective planning for vascular access (VA) is essential for initiating hemodialysis in patients with end-stage renal disease (ESRD). Currently, over 80% of patients begin treatment using a temporary access, which increases the risk of infection, complications, and mortality. Deciding when and which type of permanent access to initiate is challenging due to patient heterogeneity and uncertainty in access outcomes.

This study presents a finite-horizon Markov Decision Process (MDP) model to identify the optimal timing and choice of vascular access—arteriovenous fistula (AVF), arteriovenous graft (AVG), or wait at each epoch—for pre-dialysis patients. The model considers each patient's estimated glomerular filtration rate (eGFR), current access

status, access type, prior failures, age, and frailty. It accounts for uncertainty in access maturation and disease progression. The decision process ends when the access becomes functional or after repeated failures. The objective is to maximize expected discounted lifetime health outcomes.

We stratify the model by patient frailty and age to enable personalized decision support. Results demonstrate that optimal access initiation policies are sensitive to patient-specific characteristics and outperform heuristic approaches that follow static “fistula first” guidelines. This framework offers a tractable, policy-relevant tool for sequential medical decision-making under uncertainty, supporting personalized treatment recommendations.

3 - Optimizing Ovarian Cancer Early Detection Using Stochastic Modeling

Sayaka Ishizawa, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, Maryam Eghbalizarch, Renu S Nargund, Andrew Schaefer, M. Soheil Hemmati, Seyyed Mostafa Mousavi Janbeh Sarayi, Karen Lu, Sharon H Giordano, Larissa A Meyer, Iakovos Toumazis

Background: Ovarian cancer is the second leading cause of death among all gynecologic cancers. Yet, no effective prevention program currently exists. Well-validated natural history models (NHMs) provide insights on the natural progression of diseases and can be used to assess the effectiveness of potential interventions in silico. Currently, no histology-specific NHM exist for ovarian cancer.

Study Design: We developed histology-specific NHMs for the seven most common ovarian cancer subtypes using discrete-time Markov models. We calibrated histology-specific model parameters using Adaptive Simulated Annealing such that model outcomes reliably reproduced the observations from the Surveillance, Epidemiology, and End Results (SEER) program. We assessed goodness-of-fit using the weighted sum of root mean squared error (RMSE). We validated the NHMs on data from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) and the United Kingdom Collaborative Trial on Ovarian Cancer Screening (UKCTOCS). Moreover, the stage distribution at diagnosis was validated on PLCO data. Differences between observations and model estimates were evaluated using the Pearson’s chi-squared test.

Results: The calibrated NHM showed satisfactory goodness of fit to the SEER data, within 2% of RMSE. In model validation, none of our model estimates showed significant difference on UKCTOCS (p-value on incidence=0.399, death=0.377) or PLCO (p-value on incidence=0.207, death=0.513, stage distribution=0.498) observations.

Conclusion: The developed NHM accurately describes the histology-specific natural progression of ovarian cancer. This model could be used to assess the impact of future and emerging ovarian cancer preventive interventions, thus assisting policy makers developed effective prevention programs.

SB12

Building A Level 3 A314

Resilience and Crisis Response

Job Market Showcase

Job Market Showcase

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

1 - Generative Simulation: AI-Driven Predictive Modeling for Decision Support in Crisis Operations

Seung ho woo, Purdue University, Lafayette, IN, United States, Young-Jun Son

Crisis operations demand rapid decision-making in environments characterized by uncertainty, complexity, and data scarcity. Traditional simulation approaches often fall short in these contexts due to their reliance on extensive historical data and rigid model assumptions. This work introduces a novel paradigm, Generative Simulation, that leverages advanced artificial intelligence to create adaptive, knowledge-driven predictive models for crisis decision support. The framework integrates generative AI models with agent-based simulation to enable zero-shot scenario generation and adaptive agent behavior without requiring extensive predefined models or training data.

Our research makes three primary contributions including a zero-shot generative simulation architecture that autonomously constructs crisis scenarios from minimal inputs, LLM-embodied agents capable of realistic decision-making that adapts to evolving crisis conditions, and empirical validation comparing generative agent simulation against traditional approaches across multiple crisis domains. Experiments in disaster management, power grid failure response, and emergency resource allocation demonstrate that generative simulation achieves comparable or superior predictive accuracy while requiring significantly less domain-specific data engineering.

The implications extend beyond technical innovation by reducing the data and expertise barriers to sophisticated simulation, this approach democratizes access to advanced decision support tools. Communities with limited resources can leverage generative simulation to model complex crisis scenarios previously beyond their analytical reach. This research represents a fundamental shift in how simulation science approaches predictive modeling under uncertainty, with applications in disaster resilience, infrastructure protection, and humanitarian response operations.

2 - Machine Learning for Crisis Operations: Detecting Patterns of Vulnerability and Predicting Demand

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

Crisis environments create dynamic and uncertain operational conditions, requiring fast, data-driven responses. This research investigates how machine learning (ML) algorithms can uncover hidden patterns in unstructured data and support real-time decision-making in humanitarian operations. We apply natural language processing (NLP) to identify suspicious recruitment patterns indicative of human trafficking, using multilingual social media (Telegram) and messaging data. Beyond detection, we extend our methods to operational tasks such as demand prediction and vulnerability estimating needs for shelter, transportation, or medical aid based on textual context from digital platforms. Our approach uses transformer-based algorithms using ensemble techniques to improve accuracy across tasks. We develop a model pipeline that can generalize across regions and crises by addressing key challenges like data imbalance, language diversity, and informal communication formats. This work contributes to building adaptive, pattern-based ML systems that enhance situational awareness and resource allocation in disaster response and humanitarian logistics.

3 - When Every Second Matters: Dispatching Drones After a Disaster to Locate Victims

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

The use of drones in humanitarian and disaster management operations has gained increasing attention due to their speed and reliability in data gathering and assessments, especially where road infrastructure is disrupted. Drones reduce risks for search and rescue teams by enabling efficient victim detection without exposing responders to hazardous environments. This study presents a two-phase optimization framework for drone-based search and rescue after major disasters.

In the first phase, K-means clustering identifies potential hub locations, followed by a heuristic approach to optimize hub selection and drone allocation, ensuring inter-hub connectivity within flight range constraints. The second phase formulates a team orienteering problem to design efficient drone routing strategies, maximizing coverage and search effectiveness.

To address disaster-related uncertainty and infrastructure disruptions, the model incorporates potential facility failures and integrates alternative hubs to enhance resiliency. These hubs enable on-time replanning of operations through real-time decision-making and contingency strategies, preserving connectivity when primary hubs are compromised.

A case study in San Fernando Valley, Los Angeles, incorporates earthquake-specific data and building density. Three operational policies are evaluated that offer valuable trade-offs for decision makers: unconstrained coverage (maximal utility), threshold-based coverage (equity), and maximal covering hub selection. Results show that inter-

connectivity and alternative hubs significantly improve performance, mitigate battery and range constraints, and extend operational endurance. The findings offer insights into resilient and efficient drone deployment for post-disaster response.

SB13

Building A Level 3 A315

Queueing Systems

Job Market Showcase

Job Market Showcase

Chair: Deniz Simsek, Northwestern University, Evanston, IL, United States

1 - Join-the-Shortest Queue with Abandonment

prakirt jhunjunwala, Columbia Business School, Westbury, NY, United States

We consider a load balancing system comprised of a fixed number of single server queues, operating under the well-known Join-the-Shortest Queue policy, and where jobs/customers are impatient and abandon if they do not receive service after some (random) amount of time. In this setting, we characterize the centered and appropriately scaled steady-state queue length distribution (hereafter referred to as limiting distribution) in the limit as the abandonment rate goes to zero at the same time as the load either converges to one or is larger than one.

Depending on the arrival, service, and abandonment rates, we observe three different regimes of operation that yield three different limiting distributions. The first regime is when the system is underloaded, and its load converges relatively slowly to one. In this case, abandonments do not affect the limiting distribution, and we obtain the same exponential distribution as in the system without abandonments. When the load converges to one faster, we have the second regime, where abandonments become significant. Here, the system undergoes a phase transition, and the limiting distribution is a truncated Gaussian. Further, the third regime is when the system is heavily overloaded, and so the queue lengths are very large. In this case, we show the limiting distribution converges to a normal distribution.

2 - Socially Optimal Pricing of Downtown Parking Through the Lens of Cruising Delays

Xinyu Liu, University of Michigan, Ann Arbor, MI, United States, Vijay Subramanian, Yafeng Yin

Cruising is the search for parking spaces when a driver finds no vacancy near the desired parking location. This unpleasant experience of uncertainty and frustration is frequently encountered in downtown areas, aggravating the urban congestion and emission. Pricing has been long promoted as an efficient control to manage downtown parking, reduce cruising for parking, and enhance social welfare. Donald Shoup advocated parking pricing to meet a target occupancy of 85%, which has been widely accepted and implemented in demand-responsive parking pricing projects like SFpark project (launched by the San Francisco Municipal Transportation Authority in 2011). Despite its widespread acceptance, our literature search reveals that few studies have systematically examine the optimal target occupancy. Furthermore, with the advent of autonomous vehicles and their tendency to cruise as a substitution for parking, the optimal pricing and space allocation schemes with the mixed traffic is yet to be understood.

We formulate the parking-cruising system as a queueing network with price- and delay-sensitive arrivals and state-dependent retrial rates that are functions of congestion levels. We provide the stability conditions, propose a scalable approximation to the stationary distribution, show its equivalence to the exact solution in small systems, and validate its performance using a simulation. We leverage the estimated cruising delays to analytically and computationally characterize the target occupancy that optimizes the system welfare.

3 - Dynamic Transfer Policies for Parallel Queues

Jangwon Park, University of Toronto, Toronto, ON, Canada, Timothy Chan, Vahid Sarhangian

We consider the problem of load balancing in parallel queues by transferring customers between them at discrete points in time. Holding costs accrue as customers wait in the queue, while transfer decisions incur both fixed (setup) costs and variable costs that increase proportionally with the number and distance of transfers and vary by direction. Our work is motivated by inter-facility patient transfers to address imbalanced congestion and inequity in access to care during surges in hospital demand (e.g., pandemics). Analyzing the associated fluid control problem, we show that under general assumptions, including time-varying arrivals and convex holding costs, the optimal policy partitions the state-space into a well-defined *no-transfer region* and its complement, implying that transferring is optimal if and only if the system is sufficiently imbalanced. In the absence of fixed transfer costs, an optimal policy moves the state to the no-transfer region's boundary; in contrast, with fixed costs, the state is moved to its relative interior. Leveraging our structural results, we propose a simulation-based Approximate Dynamic Programming (ADP) algorithm to find good transfer policies for the stochastic system. We investigate the performance and robustness of the fluid and ADP policies in a case study calibrated using data during the pandemic in the Greater Toronto Area, which demonstrates that transferring patients between hospitals could result in up to 27.7% reduction in total cost with relatively few transfers.

4 - Large Service Systems with Patience- OR Delay-Dependent Service TIMES: Fluid Approximations, Design and Control

Deniz Simsek, Northwestern University, Evanston, IL, United States, Achal Bassamboo, Ohad Perry

We consider a single-pool many-server queueing system, assuming the service time of each customer depends on the delay of that customer in queue. Such dependence can be due to the customers having finite patience for waiting that depends on their individual service requirement, or due to having their service-time distribution be a function of the time spent in queue. We refer to the former dependence mechanism as “exogenous dependence” and to that latter as “endogenous dependence.” Since exact analysis of the stochastic system under either dependence mechanism is intractable, we propose a deterministic approximation for the (mean) queueing dynamics, referred to as a Unified Fluid Model (UFM). For constant arrival rates, we characterize conditions for the existence of a unique stationary point for the UFM and prove that those conditions always hold when the dependence is exogenous. However, the UFM may possess multiple equilibria, with each equilibrium point being either locally stable or unstable. The implication of the multi-stability of the UFM for the stochastic system is two-fold. First, the stochastic fluctuations in steady state may be an order of magnitude larger than the typical fluctuations in many-server queueing systems. Second, an underloaded or critically loaded system may experience congestion collapse, namely, it becomes severely congested. Finally, for systems with delay-dependent services, which are prone to multi-stability, we propose design and control mechanisms achieving the performance level of the most desirable equilibrium of the UFM while maintaining unique stationary behavior.

SB14

Building A Level 3 A316

Design of Incentives and Allocations

Job Market Showcase

Job Market Showcase

Chair: Eklavya Sharma, ISE, UIUC, Urbana, IL, United States

1 - Bayesian Mechanism Design for Blockchain Transaction Fee Allocation

Zishuo Zhao, University of Illinois Urbana-Champaign, Urbana, IL, United States, Xi Chen, David Simchi-Levi, Yuan Zhou

In blockchain systems, the design of transaction fee mechanisms (TFMs) is essential for stability and satisfaction for both miners and users. A recent work has proven the impossibility of collusion-proof mechanisms that achieve both nonzero miner revenue and Dominant Strategy Incentive Compatibility (DSIC) for users. However, a positive miner revenue is important in practice to motivate miners. To address this challenge, we consider a *Bayesian game* setting and relax the DSIC requirement for users to Bayesian Nash Incentive Compatibility (BNIC). In particular, we propose an auxiliary mechanism method that makes connections between BNIC and DSIC

mechanisms. With the auxiliary mechanism method, we design a TFM based on the multinomial logit (MNL) choice model, and prove that the TFM has both BNIC and collusion-proof properties with an asymptotic constant-factor approximation of optimal miner revenue for i.i.d. bounded valuations. Our result breaks the zero-revenue barrier while preserving truthfulness and collusion-proof properties.

2 - Single-Sample and Robust Online Resource Allocation

Yifan Wang, Georgia Institute of Technology, Atlanta, GA, United States, Rohan Ghuge, Sahil Singla

Online Resource Allocation problem is a central problem in many areas of Computer Science, Operations Research, and Economics. In this problem, we sequentially receive n stochastic requests for m kinds of shared resources. Each request can be satisfied in multiple ways, consuming different amounts of resources and generating different values. The goal is to achieve a $(1 - \epsilon)$ -approximation to the hindsight optimum, where $\epsilon > 0$ is a small constant, assuming each resource has a large budget.

We investigate the learnability and robustness of online resource allocation. Our primary contribution is a novel Exponential Pricing algorithm with the following properties:

1. It requires only a single sample from each of the n request distributions to achieve a $(1 - \epsilon)$ -approximation for online resource allocation with large budgets. Such an algorithm was previously unknown, even with access to polynomially many samples, as prior works either assumed full distributional knowledge or was limited to i.i.d. or random-order arrivals.
2. It is robust to corruptions in the outliers model and the value augmentation model.
3. It operates as a simple item-pricing algorithm that ensures incentive compatibility.

The intuition behind Exponential Pricing algorithm is that the resource prices should adjust exponentially as it is overused or underused. It differs from conventional approaches that use an online learning algorithm for item pricing. This departure avoids running out of resources entirely, but loses the usual no-regret properties of online learning algorithms, necessitating a new analytical approach.

3 - Mission vs. Profit: The Interplay Between Social Service Agencies and Private Service Providers

Gulten Busra Karkili, University of Massachusetts-Amherst, Amherst, MA, United States, Senay Solak

Voucher-based subsidy programs are a key policy tool for improving access to essential services such as childcare and housing, but their success depends heavily on voluntary participation from private providers who face uncertainty and financial trade-offs. This paper presents an optimization framework capturing the strategic interaction between a mission-oriented agency and quality-differentiated providers within a voucher system. The agency invests in providers to increase service quality and encourage participation, while providers decide whether to participate and what portion of their capacity they expect to be used by voucher beneficiaries, balancing uncertain private-payer demand and capacity utilization risk.

We characterize how provider participation and planned participation levels respond to changes in agency investment, market competition, and provider characteristics. Our analysis reveals that under certain conditions, providers can be incentivized to participate even without direct funding, either through increased voucher payments or indirect competitive effects. We also identify operational scenarios where higher investments paradoxically lead to non-participation by service providers.

We apply the model using data from Massachusetts' C3 childcare grant program. Results show that optimized investments can increase service access by over 10% without additional funding, demonstrating how targeted financial strategies can significantly improve program outcomes. This work provides actionable insights for agencies designing more effective and efficient public-private partnerships in essential service delivery.

4 - New Fairness Concepts for Allocating Indivisible Items

Eklavya Sharma, ISE, UIUC, Urbana, IL, United States, Ioannis Caragiannis, Jugal Garg, Nidhi Rath, Giovanna Varricchio

For the fundamental problem of fairly dividing a set of indivisible items among agents, envy-freeness up to any item (EFX) and maximin-share fairness (MMS) are arguably the most compelling fairness concepts proposed until now. Unfortunately, despite significant efforts over the past few years, whether EFX allocations always exist is still an enigmatic open problem, let alone their efficient computation. Furthermore, today we know that MMS allocations are not always guaranteed to exist. These facts weaken the usefulness of both EFX and MMS, albeit their appealing conceptual characteristics. We propose two alternative fairness concepts, called epistemic EFX (EEFX) and minimum EFX share fairness (MXS), inspired by EFX and MMS. For both, we explore their relationships to well-studied fairness notions and, more importantly, prove that EEFX and MXS allocations always exist and can be computed efficiently for additive valuations. Our results justify that the new fairness concepts can be excellent alternatives to EFX and MMS.

SB15

Building A Level 4 A401

ATS Best Student Competition

Award Session

Air Transportation Section

Chair: Haochen Wu, University of Michigan, Ann Arbor, Ann Arbor, MI, United States

Co-Chair: Minghao Chen, Columbia University, New York, United States

1 - Integrating Advanced Air Mobility into Agent-Based Demand Modeling via Adaptive Agent Behavior

Mahdi Monemi Nodehi, University of South Florida, Tampa, FL, United States

coming soon

2 - Air Carrier Capacity Allocation for Collaborative Humanitarian Logistics

Micah Borrero, University of Michigan, Ann Arbor, MI, United States

Vulnerable communities are increasingly exposed to natural disasters due to global climate change. Coordinated humanitarian responses to impacted regions are often logistically complex and expensive. Given the high cost and potential infeasibility of a well-coordinated post-disaster response, the preemptive, strategic placement of humanitarian logistics assets could yield numerous benefits, including lowering potential operational costs, improving response times, and ensuring maximum coverage of disaster-impacted communities. We propose a bi-level optimization architecture in which the leader-level problem models the interaction between fleet operators and a logistics agency, while the follower-level problem models the interaction between the logistics agency and downstream aid delivery operations. Specifically, the leader model captures a fleet operator's decision process when weighing the cost of reserving air freight capacity against the risk of unmet humanitarian demand. The follower model subsequently responds to realized scenarios by routing aid through a potential multimodal air-ground network, combining reserved air capacity with pre-positioned trucks to minimize delivery time. Preliminary results from sensitivity analyses reveal tradeoffs between operator costs and humanitarian outcomes. These show how organizational priorities influence capacity reservation strategies and identify conditions where under- or over-reserving capacity becomes particularly costly. Ongoing work incorporates operational uncertainties, such as ad-hoc payload adjustments by airlines, that can significantly disrupt time-sensitive missions and historical air carrier data. We aim to develop robust optimization models that yield resilient capacity allocation strategies, strengthening the reliability of humanitarian supply chains under growing global uncertainty.

3 - Analyzing Inter-Airline Cooperative Coordination Mechanisms Under A Time-Varying Cost Function

Tolga Atabas, University of Michigan, Ann Arbor, MI, United States

Airlines frequently face delays from disruptions that reduce airport capacity, such as adverse weather. To balance constrained capacity with demand, air traffic managers reschedule flights, often causing widespread delays. Airlines model delay costs for each flight as a function of delay severity (e.g., hours) and prior work on intra- and

inter-airline slot swapping has applied such functions to minimize systemwide delay costs. However, an existing inter-airline swapping mechanism—the Delay Ledger (DELED), which uses a ledger-based system to coordinate scheduling decisions among airlines—assumes fixed-rate cost functions. This assumption fails to capture real-world cost surges arising from missed connections, crew disruptions, and tail-propagated delays.

This study introduces a convex, stochastically generated piecewise linear cost function that captures the time-varying nature of delay costs in the real world. We model surge events with a Poisson process and integrate the resulting functions into the DELED framework. This enhancement improves both the realism and robustness of delay management models.

Our results show that incorporating time-varying cost structures yields more efficient delay assignments compared to fixed-rate assumptions. Numerical experiments further demonstrate that DELED's ledger-based coordination outperforms a round-robin rotational mechanism in allocating delay across airlines. Finally, we conduct a sensitivity analysis to examine how parameter changes affect DELED's ability to reduce system-level delay costs, highlighting the framework's adaptability across diverse operating conditions.

4 - A Dynamic Model for Airline Fleeting and Scheduling

Jingxing Wang, University of California, Berkeley, Berkeley, CA, United States

The global airline industry slumped during COVID-19: demand was volatile, cancellations surged, and airlines struggled to align capacity with shifting demand, pushing load factors to historic lows. In the recovery, airlines have institutionalized several measures first adopted during the pandemic, such as offering passengers greater flexibility in booking and cancellation. As a result, airlines are increasingly required to be agile and adaptive in their operational decisions. We introduce a dynamic model to help airlines make adaptive fleeting and scheduling decisions based on stochastically evolving bookings. The model incorporates both arrivals (new bookings) and departures (cancellations). This significantly generalizes and strengthens previous modeling attempts. We develop a Lagrangian relaxation framework that decomposes the dynamic program on a large time-space network into separable flight-level problems, and we analyze the tightness of this relaxation under representative flight network structures. Owing to the complexities of this network, our Lagrangian dual problem is not straightforward to solve. We therefore develop a tailored yet simple projected subgradient algorithm that exploits the structure of this network for efficient solutions. We establish the correctness and convergence properties of this procedure. Our analysis yields new theoretical, algorithmic, and managerial insights into the dynamic fleeting and scheduling problem. We present computational experiments based on real-world airline booking data to demonstrate the potential benefits of this approach.

5 - Aircraft Trajectory Optimization for Combinatorial Sampling in the Airspace

6 - Penalty Functions for Truck-and-Drone Routing Problem

Ke Xu, University of Southern California, Los Angeles, CA, United States

Drones are increasingly being integrated into last-mile delivery networks to alleviate reliance on conventional ground trucks. This collaboration can involve various combinations of trucks and drones and incorporate diverse operational constraints tailored to specific delivery requirements. In previous studies, however, each problem variant is typically addressed by a specialized solution method. In this study, we introduce a versatile heuristic approach based on the Lin-Kernighan-Helsgaun algorithm with penalty functions to solve a wide range of drone-assisted delivery variants with a unified solution framework. Its flexibility and effectiveness are demonstrated through numerical experiments on benchmark instances designed for several representative problem variants, including the Flying Sidekick Traveling Salesman Problem (FSTSP), the Traveling Salesman Problem with Multiple Drones (TSP-MD) and the Vehicle Routing Problem with Drones (VRPD).

7 - Optimization Model for Large-Scale Long-Term Aircraft Maintenance Scheduling and Station Assignment

JohnPaul Adimonyemma, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Yanshuo Sun

The primary focus of the existing aircraft maintenance optimization literature is on short-term aircraft maintenance routing. By contrast, only a few studies address long-term maintenance planning, which predominantly emphasize scheduling decisions: none of them have examined the critical aspect of assigning aircraft to maintenance stations, a key challenge for major airlines with a network of maintenance facilities. Additional research gaps include (1) insufficient consideration of practical maintenance-related constraints (e.g., station access limit and aircraft rotation requirement) and (2) the absence of scalable algorithms capable of solving real-world problem instances involving hundreds of aircraft and dozens of maintenance stations. To fill those gaps, this study introduces a joint optimization model for aircraft maintenance scheduling and station assignment to assist a U.S.-based airline in creating long-term plans for a fleet of over 800 aircraft of multiple subfleet types. To address the computational challenges, we propose two decomposition approaches: one based on decision types and the other on time horizons. Extensive computational experiments using real-world data from the collaborating airline demonstrate that these approaches can reduce computation time by up to 80% with a minimal increase in the optimization objective (less than 2%). The proposed model is expected to streamline maintenance planning efforts and enhance outcomes for aircraft maintenance professionals.

8 - Detecting Anomalous Bird Flock Patterns Using Coarse-Grained Complexity: Toward Safer UAS Operations Near Airports

9 - Optimizing Multi Team Agricultural Drone Deployment and Routing with Weather Aware Planning

Ethan Kolby, University of Michigan, Ann Arbor, MI, United States

UAS can service dozens to hundreds of fields on a county-scale, but real operations require routing that adapts to fast-changing weather while coordinating multiple teams. This work extends a prior facility-location→vehicle-routing (FLP→VRP) pipeline by making deployments weather-aware. For each rolling forecast horizon, gridded wind and precipitation products are interpolated to a continuous vector field when needed, thresholds are applied to polygonize exceedance regions, and a facility-time availability matrix is built. Availability enters the optimization as time-windowed closures: closed sites cannot open or dispatch during affected windows; open sites remain routable. The regional problem proceeds as FLP→VRP: fields are assigned to feasible facilities, and shortest tours over fields and required facility legs are computed for each team. As forecasts update, the solver re-optimizes, shifting workload to open sites and rerouting around transient closures with minimal manual intervention. We outline a practical ingest-to-optimization loop: (1) read forecast data; (2) interpolate sparse fields; (3) delineate exceedance polygons; (4) label facility-time pairs open/closed; and (5) pass the availability mask into the FLP→VRP solver for rolling-horizon updates. Ongoing work automates the weather ingest and interpolation, incorporates scenario-based wind uncertainty to harden availability labels, and evaluates regional service levels under realistic, time-varying closures. The result moves multi-team UAS operations toward robust, forecast-integrated planning and execution.

10 - Towards Efficient Aircraft Turnaround: a Mesoscopic Model for Equipment Routing and Scheduling

Marcus Koh, Singapore University of Technology and Design, Singapore, Singapore

Efficient aircraft turnaround relies on tightly coordinated ground operations carried out with limited Ground Service Equipment (GSE). As air traffic continues to grow, pressure on GSE allocation and scheduling intensifies, with poor management risking delays, higher fuel consumption, reduced passenger satisfaction, and even cancellations. To address this, we frame turnaround operations as a Vehicle Routing Problem (VRP), extending beyond classical VRP models by integrating aviation-specific complexities such as heterogeneous fleets, multiple task types, precedence constraints, reloading needs, and strict time windows.

We propose a decomposition approach where the turnaround process is structured into four interrelated subproblems—baggage operations, waste and water services, passenger-dependent operations, and aircraft pushbacks—each modelled and solved individually before integration into a master problem. Among these, the baggage subproblem is prioritized as the foundation, given its operational intricacy and strong interdependencies between the GSEs. By tackling baggage operations first, we thus establish a generalizable modelling framework that can be further extended to capture the full turnaround process.

11 - Bandit-Enabled Dynamic Vehicle Dispatch and Routing in Hazardous Environments

Jimin Choi, University of Michigan, Ann Arbor, MI, United States

Hazardous environments such as chemical spills, radiological zones, and bio-contaminated sites pose significant threats to human safety and public infrastructure. Rapid and reliable hazard mitigation in these settings often unsafe for humans, calling for autonomous systems that can adaptively sense and respond to evolving risks. This paper presents a decision-making framework for autonomous vehicle dispatch in hazardous environments with uncertain and evolving risk levels. The system integrates a Bayesian Upper Confidence Bound (UCB) sensing strategy with task-specific vehicle routing problems with profits (VRPP), enabling adaptive coordination of unmanned aerial vehicles (UAVs) for hazard sensing and unmanned ground vehicles (UGVs) for cleaning. Using VRPP allows selective site visits under resource constraints by assigning each site a visit value that reflects sensing or cleaning priorities. Site-level hazard beliefs are maintained through a time-weighted Bayesian update. UCB scores guide UAV routing to balance exploration and exploitation under uncertainty, while UGV routes are optimized to maximize expected hazard

reduction under resource constraints. Simulation results demonstrate that our framework reduces the number of dispatch cycles to resolve hazards by around 30% on average compared to baseline dispatch strategies, underscoring the value of uncertainty-aware vehicle dispatch for reliable hazard mitigation.

SB16

Building A Level 4 A402

AI-enabled Decision Intelligence in Business Operations

Invited Session

The Practice Section of Informs

Chair: Jiaxi Zhu, Google, Mountain View, 94040, United States

Co-Chair: Bhubalan Mani, Garmin International, Inc., 13885 W 153RD TER, Olathe, KS, 66062, United States

1 - From Complexity to Clarity: AI in Action for Resilient and Efficient Supply Chains

Bhubalan Mani, Garmin International, Inc., Olathe, KS, United States

The current global economy with its unpredictable nature and interconnected systems puts supply chains under continuous stress from unexpected disruptions, changing market demands and internal planning system inefficiencies. The ability of traditional rule-based systems to handle uncertainties is limited. The presentation “From Complexity to Clarity: AI in Action for Resilient and Efficient Supply Chains” demonstrates how AI and analytics solutions solve real business problems to create better decision-making processes. My two decades of experience working with enterprise supply chain systems and analytics allows me to share practical lessons from recent projects that integrated AI and machine learning into ERP-driven planning functions to optimize safety stock levels and balance inventory across multiple echelons and forecast demand under uncertain conditions. The implemented initiatives delivered better operational performance while moving supply networks toward predictive decision-making at a global level. The session will demonstrate how organizations can design and implement AI-enabled planning tools without having to overhaul existing systems by strategically layering intelligence over operational data. I will also discuss critical success factors, including cross-functional collaboration, data readiness, and change management. The participants will have a clear understanding of how to turn AI from a buzzword into a practical enabler of supply chain resilience and efficiency. This session will be useful for practitioners who want to know how to implement AI and decision makers who want to know if AI is a good fit for their operations.

2 - AI-Enabled Customer Sentiment Analysis in B2b Sales Operations

Jiaxi Zhu, Google, Mountain View, CA, United States

In complex B2B environments, sales effectiveness often relies on qualitative signals that are difficult to quantify or scale. This talk presents a case study on how AI-enabled sentiment analysis was operationalized within a large-scale B2B sales organization to improve customer engagement. We explore a hybrid framework that combines natural language processing (NLP) and large language models (LLMs) with structured sales pipeline data to surface sentiment trends, customer risk indicators, and emerging objections in real time. The system was deployed across global sales operations, with models trained on call transcripts and was integrated directly into frontline decision workflows. Key outcomes include improved deal prioritization and accelerated resolution of customer concerns. We also highlight lessons learned in implementation, including cross-functional buy-in, model interpretability, and feedback loops between AI outputs and sales strategy. This presentation offers practical guidance for analytics leaders and OR professionals working at the intersection of AI, sales, and customer insights.

3 - From Complexity to Clarity: AI in Action for Resilient and Efficient Supply Chains

The current global economy with its unpredictable nature and interconnected systems puts supply chains under continuous stress from unexpected disruptions, changing market demands and internal planning system inefficiencies. The ability of traditional rule-based systems to handle uncertainties is limited. The presentation “From Complexity to Clarity: AI in Action for Resilient and Efficient Supply Chains” demonstrates how AI and analytics solutions solve real business problems to create better decision-making processes. My two decades of experience working with enterprise supply chain systems and analytics allows me to share practical lessons from recent projects that integrated AI and machine learning into ERP-driven planning functions to optimize safety stock levels and balance inventory across multiple echelons and forecast demand under uncertain conditions. The implemented initiatives delivered better operational performance while moving supply networks toward predictive decision-making at a global level. The session will demonstrate how organizations can design and implement AI-enabled planning tools without having to overhaul existing systems by strategically layering intelligence over operational data. I will also discuss critical success factors, including cross-functional collaboration, data readiness, and change management. The participants will have a clear understanding of how to turn AI from a buzzword into a practical enabler of supply chain resilience and efficiency. This session will be useful for practitioners who want to know how to implement AI and decision makers who want to know if AI is a good fit for their operations.

SB17

Building A Level 4 A403

Inventory Management at Amazon

Invited Session

Practice Track

Chair: Chun Ye, Amazon.com, Seattle, WA, 98101, United States

Co-Chair: Ishan Bansal, Amazon, Lynnwood, WA, United States

1 - Managing Inventory Flows in a Multi-Echelon Network - a Simulation Study

Zihao Li, Amazon, Bellevue, WA, United States

Inbound Network Simulation (INSim) is a Discrete Event Simulation (DES) tool designed to replicate and experiment with various inbound network structures and inventory control strategies. In a multi-echelon inbound network, inventory is received, stored, and distributed to different facilities. We demonstrate how INSim can be used to experiment with different inventory control strategies.

2 - Inventory Placement Planning - Overcoming the Limitations of Product Grouping

Ishan Bansal, Amazon, Bellevue, WA, United States

Online retailers face the critical challenge of optimally distributing inventory across fulfillment centers under demand uncertainty. While traditional approaches allocate each product independently based on regional demand patterns, ship and store capacity constraints create interdependencies across product placement decisions. Given extensive product catalogs, retailers commonly group similar products to make optimization problems tractable.

However, this product grouping introduces significant approximation errors, particularly for tail products with sparse demand. Existing literature often treats products within a group as fungible, resulting in overly optimistic placement strategies for low-demand items. This talk presents novel techniques to address these limitations and introduces an algorithm that effectively disaggregates group-level placement solutions into precise product-level allocation decisions. Our approach maintains computational efficiency while significantly improving inventory positioning accuracy for the full product spectrum.

3 - Optimizing Inventory Placement in Speed-Sensitive Fulfillment Networks with Consumer Substitution Behavior

Katja Meuche, Georgia Institute of Technology, Atlanta, GA, United States, Cristiana Lara, Louis Faugere, Benoit Montreuil

The traditional Newsvendor model, while analytically elegant, struggles with modern e-commerce fulfillment networks where consumer-driven substitution occurs due to stock-outs and long delivery times. This research develops a two-stage stochastic optimization framework that incorporates consumer substitution behavior for inventory

decisions across multiple fulfillment locations.

We extend the classical Newsvendor model by including both stockout-based and delivery speed-based substitution behaviors. Our model optimizes inventory placement across a fulfillment network, considering product substitutability, delivery time expectations, and capacity constraints to reduce operating costs.

Recognizing that temporal demand sequences drive real-world operations, we explore how to approximate consumer-driven substitution behavior through various constraint formulations and a comparative simulation study. By capturing the dynamics of consumer substitution in multi-location networks, we assess the impact of substitution-aware inventory placement on expected costs, service levels, and inventory levels.

Our preliminary experiments show that substitution-aware inventory placement can reduce inventory levels by 14% on average while maintaining service targets. This research enhances both the theoretical understanding of inventory models and their practical applications in retail. By accurately representing consumer substitution behavior, we help retailers manage the trade-off between costs and service levels in competitive omni-channel environments.

SB18

Building A Level 4 A404

Transportation Safety, Security, and Sustainability

Invited Session

TSL: Intelligent Transportation Systems

Chair: Haitao Li, University of Missouri - St. Louis, Saint Louis, MO, United States

1 - Driving Renewable Integration: Dual-use Carshare Fleets for Mobility and Grid Services

Kevin Freymiller, Carnegie Mellon University, Pittsburgh, PA, United States, Haocheng Duan, Sean Qian

As institutions increasingly deploy renewable energy generation, they typically seek to reduce their energy costs and carbon emissions. Solar energy deployments continue to grow worldwide and so does the need for local energy storage alongside it. Fixed battery storage systems remain too expensive in most cases to provide significant savings on energy. Institutions also generate significant transportation demand, and many are transitioning to battery electric vehicles to meet these needs. These batteries can serve a dual use for energy storage and transportation when operated optimally, and can even serve to transport energy between multiple locations where energy costs and emissions vary, providing an additional benefit to the system. We investigate how an electrified carshare fleet providing mobility services within an institutional campus can provide mutual benefits beyond transportation by storing and moving energy in the vehicle battery. By utilizing vehicle-to-grid technology, vehicles charge and discharge as needed at different sites depending on site specific energy prices, generation, and demand, reducing energy costs, and increasing utilization of renewable energy. We show that charging vehicles at sites with excess solar generation reduces cost of energy for transportation and also sites where V2G delivers energy to offset building demand. This approach shows benefit increases when energy prices are highest, which coincides with increased travel delays. When energy costs are low and transporting energy has less benefit, travel delays are decreased. Thus we show that considering tradeoffs between the system leads to an overall systemwide benefit.

2 - Sensor-Enabled Intelligent Distribution for Perishable Food Products

Haitao Li, University of Missouri - St. Louis, Saint Louis, MO, United States

With the technology of rapid pathogen sensors being deployed throughout an end-to-end poultry supply chain, *Salmonella* testing results can be obtained and shared in real time, which makes it possible to improve processing, distribution and transportation operations. We develop a mixed-integer programming (MIP) model to automate and optimize the diversion decision in poultry processing, distribution of limited supply to the markets in need, and cold chain options, e.g., transportation modes, temperature setting, while satisfying supply capacity, market demand, and food safety standards. The novelty of our model is its ability to model diversion decision in poultry processing, and the consideration of pathogen growth and decrease of shelf-time and product value during cold chain transportation. A case study of an end-to-end poultry supply chain is conducted to show the type of decision support that can be provided by our model and how it simultaneously addresses food safety and operating efficiency.

3 - Energy and Emission Efficient Speed Suggestion System for Human Drivers: Simulation and Real-World Validation

Maziar Zamanpour, University of Minnesota, Minneapolis, MN, United States

Advancements in vehicle connectivity and automation have presented numerous opportunities in transportation, such as increased throughput, enhanced safety, and improved energy efficiency. While connected and automated vehicles (CAVs) offer significant energy savings, their widespread implementation faces substantial practical challenges. This study presents an energy-efficient speed suggestion system for human-driven connected vehicles (ESSCV), bridging the gap between conventional vehicles and fully autonomous ones. A Model Predictive Control (MPC), commonly used in eco-driving for CAVs, calculates optimal acceleration, providing speed advisories to human drivers. The driver's response to these advisories is modeled using an adaptive Intelligent Driver Model (IDM) calibrated in real-time. To validate the simulation results, a real-world road test was conducted in Minnesota using a connected vehicle communicating with a traffic signal. Initially, the system displayed speed suggestions to the driver, and the IDM was trained based on observed driver responses. Subsequently, an actual road test was performed, demonstrating practical feasibility and effectiveness. The road test confirmed that drivers could reliably follow the advisories, achieving energy savings consistent with simulation predictions. Overall, the integrated ESSCV framework, combining traffic prediction, MPC-based optimization, and real-time driver behavior modeling, achieves an average energy saving of up to 12% compared to conventional car-following behavior at signalized intersections, with minimal travel time impacts. This study demonstrates the practical viability and benefits of integrating ESSCV into existing speed control mechanisms, providing a cost-effective interim solution as the technology for fully automated vehicles continues to evolve.

SB19

Building A Level 4 A405

Applied Probability in Algorithmic and Operational Networks II

Invited Session

Applied Probability Society

Chair: Sophie Yu, The Wharton School of the University of Pennsylvania, Philadelphia, PA, United States

Co-Chair: Yeganeh Alimohammadi, USC, Los Angeles, CA, United States

1 - A Random Model of Supply Chain Networks

Georgina Hall, INSEAD, Fontainebleau, France, Philippe Blaettchen, Andre Calmon

Supply chain problems are frequently formulated as optimization problems over graphs representing complex networks of interlinked input-output relationships. Frequently, these problems are hard, so researchers rely on analyzing stylized structures or developing heuristic solutions. Yet, a scarcity of real-world data has hindered our understanding of how these exact and heuristic solutions perform in practice and whether managerial insights carry over from these simplified settings. We address this critical gap by introducing RG4SC, a versatile random graph model for creating "test tracks" for supply chain management research. RG4SC's simple micro-foundations and interpretable input parameters allow for systematically generating diverse and realistic network structures. We demonstrate its empirical validity and that it more adequately represents real-world supply chain networks than existing random models. We then showcase RG4SC's utility for research through a case study on the Guaranteed Service Model, a widely used framework for safety stock optimization. Our analysis shows how RG4SC can be central to uncovering novel managerial insights, analyzing the computational complexity of algorithms, benchmarking heuristics, and training machine learning models.

2 - Two-Sided Flexibility in Platforms

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

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

3 - Aligning Multiple Inhomogeneous Random Graphs: Fundamental Limits of Exact Recovery

Taha Ameen, University of Illinois Urbana-Champaign, Urbana, IL, United States, Bruce Hajek

We explore impacts of network heterogeneity and multiple networks on the graph matching problem, i.e. to recover the underlying correspondence between correlated unlabeled graphs. This problem captures essential challenges in operations: matching users across platforms for targeted advertising, aligning supply chains for merger analysis, correlating sensor networks for infrastructure monitoring, or aligning feature representations across different datasets for machine learning. The insight is that networks exhibit correlation (e.g. two friends on Facebook likely also follow each other on Twitter), but the correlation is obscured by unknown node correspondences.

Real-world networks have dense clusters, spatial correlations, and high degree superstar nodes. How do these factors influence thresholds for graph matching? Using inhomogeneous random graph models, we characterize phase transitions for exact recovery and show that heterogeneity can both help and hinder matching performance depending on the correlation structure, providing insight into which network features are most important for inference. Our multi-graph matching algorithm first partially matches graphs pairwise and then combines the partial matchings by transitivity. Its analysis yields a sufficient condition on the problem parameters to exactly match all nodes across all the graphs. In the setting of homogeneous (Erdős-Rényi) graphs, we show that this condition is also necessary, i.e. the algorithm works down to the information theoretic threshold. This reveals a scenario where exact matching between two graphs alone is impossible, but leveraging more than two graphs allows exact matching among all the graphs.

4 - The Uniformity Principle in Spatial Matching

Sophie Yu, The Wharton School of the University of Pennsylvania, Philadelphia, PA, United States, Taha Ameen, Flore Sentenac

Platforms matching mobile supply to spatially distributed demand face a fundamental design choice: given a fixed budget, how should service ranges be distributed among supply units? We model this via bipartite random geometric graphs with n supply and m demand nodes uniformly distributed on $[0,1]^k$ ($k \geq 1$), where an edge between a supply and demand node forms when the demand node is within the supply node's allocated service range. We seek to maximize the expected size of the maximum matching under a fixed total range budget. Our main result establishes that uniform allocation—equal ranges for all supply units—strictly dominates heterogeneous allocation for all dimensions $k \geq 1$. This dominance stems from diminishing marginal returns to range expansion and limited interference from bounded ranges fragmenting the graph. Additionally, for $k=1$, we characterize the size of maximum matching via Markov chain embedding and derive exact matching formulas for special cases. Our results provide theoretical justification for equalizing service territories in applications such as autonomous vehicles, drones, and emergency services.

5 - AI Supply Chains: An Emerging Ecosystem of AI Actors, Products, and Services

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

The widespread adoption of AI in recent years has led to the emergence of AI supply chains: complex networks of AI actors contributing models, datasets, and more to the development of AI products and services. AI supply chains have many implications yet are poorly understood. In this work, we take a first step toward a formal study of AI supply chains and their implications, providing two illustrative case studies indicating that both AI development and regulation are complicated in the presence of supply chains. We begin by presenting a brief historical perspective on AI supply chains, discussing how their rise reflects a longstanding shift towards specialization and outsourcing that signals the healthy growth of the AI industry. We then model AI supply chains as directed graphs and demonstrate the power of this abstraction by connecting examples of AI issues to graph properties. Finally, we examine two case studies in detail, providing theoretical and empirical results in both. In the first, we show that information passing (specifically, of explanations) along the AI supply chains is imperfect, which can result in misunderstandings that have real-world implications. In the second, we show that upstream design choices (e.g., by base model providers) have downstream consequences (e.g., on AI products fine-tuned on the base model). Together, our findings motivate further study of AI supply chains and their increasingly salient social, economic, regulatory, and technical implications.

SB20

Building A Level 4 A406

Best Paper Award - Telecommunications and Network Analytics Section

Award Session

Telecommunications and Network Analytics

Chair: Maryam Daryalal, HEC Montreal, Montreal, QC, Canada

1 - Political Districting to Minimize County Splits

Maral Shahmizad, Delta Air Lines, Atlanta, GA, United States, Austin Buchanan

When partitioning a state into political districts, a common criterion is that political subdivisions, like counties, should not be split across multiple districts. This criterion is encoded into most state constitutions and is sometimes enforced quite strictly by the courts. However, map drawers, courts, and the public typically do not know what amount of splitting is truly necessary, even to satisfy basic criteria, like contiguity and population balance. In this paper, we provide answers for all congressional, state senate, and state house districts in the United States using 2020 census data. Our approach is based on integer programming. The associated codes and experimental results are available on GitHub.

2 - The Driver-Aide Problem: Coordinated Logistics for Last-Mile Delivery

3 - Maximizing Engagement in Large-Scale Social Networks

Hamidreza Validi, Texas Tech University, Lubbock, TX, United States, Samuel Kroger, Illya Hicks

Motivated by the importance of user engagement as a crucial element in cascading leaving of users from a social network, we study identifying a largest relaxed variant of a degree-based cohesive subgraph: the maximum anchored k -core problem. Given graph $G=(V, E)$ and integers k and b , the maximum anchored k -core problem seeks to find a largest subset of vertices S that induces a subgraph with at least $|S|-b$ vertices of degree at least k . We introduce a new integer programming (IP) formulation for the maximum anchored k -core problem and conduct a polyhedral study on the polytope of the problem. We show the linear programming relaxation of the proposed IP model is at least as strong as that of a naïve formulation. We also identify facet-defining inequalities of the IP formulation. Furthermore, we develop inequalities and fixing procedures to improve the computational performance of our IP model. We use benchmark instances to compare the computational performance of the IP model with (i) the naïve IP formulation and (ii) two existing heuristic algorithms. Our proposed IP model can optimally solve half of the benchmark instances that cannot be solved to optimality either by the naïve model or the existing heuristic approaches.

SB21

Building A Level 4 A407

The Present and Future of Operational Resilience in Air Transportation

Panel Session

Transportation Science and Logistics (TSL)

Co-Chair: Alexander Estes, University of Maryland, College Park, MD, United States

1 - Moderator Panelist

Alexander Estes, University of Maryland, College Park, MD, United States

Recent, high-profile events have disrupted the national airspace system, resulting in significant delays and cancellations. Examples include winter storms in the southeast and the CrowdStrike software outage. These events demonstrate the need for resilient air transportation operations. When extreme weather, infrastructure outages, and other disruptive events degrade operations, it is critical to contain these events and avoid system collapse. This panel will discuss present and future efforts to improve operational resilience. Panelists will discuss what resilience means to their organization and the current efforts that are being taken to improve resilience. Panelists will also discuss current and future challenges, and the potential for operations researchers and management scientists to contribute to this topic.

2 - Panelist

Mark Hansen, University of California-Berkeley, Berkeley, CA, United States

SB22

Building A Level 4 A408

APS Distinguished Lecture Jiaming Xu

Invited Session

Applied Probability Society

Chair: Jiaming Xu, Duke University, Durham, NC, United States

1 - Challenges and Recent Advances in Network Inference

Jiaming Xu, Duke University, Durham, NC, United States

Inferring hidden structures in networks is a fundamental problem with wide-ranging applications in fields such as sociology, biology, and computer science. It has drawn significant attention from the applied probability community and beyond. Thanks to collective efforts over the past decade, substantial progress has been made in characterizing fundamental detection and estimation thresholds, as well as in designing efficient algorithms. This tutorial will discuss key challenges that remain, survey recent advancements addressing them, and outline promising directions for future research.

SB23

Building A Level 4 A409

Life in Academia

Panel Session

Minority Issues Forum

Co-Chair: Diana Ramirez-Rios, University at Buffalo, Buffalo, NY, United States

1 - Panelist

Jose Walteros, University at Buffalo, Buffalo, NY, United States

2 - Panelist

Deniz Besik, University of Richmond, Richmond, VA, United States

3 - Panelist

Gina Galindo, Universidad del Norte, Barranquilla, Colombia

4 - Panelist

Zhiyuan Wei, California Polytechnic State University, San Luis Obispo, San Luis Obispo, CA, United States

SB24

Building A Level 4 A410

The Role of Optimization in the Decarbonized Energy Systems of the Future

Tutorial session

Tutorials

Chair: Brian Denton, University of Michigan, Ann Arbor, MI, United States

1 - The Role of Optimization in The Decarbonized Energy Systems of The Future

Miguel Anjos, University of Edinburgh, Edinburgh, United Kingdom

Energy is a fundamental need of human activity. Electricity in particular is a critical resource for society in the 21st century and its ubiquitous use in our houses and cities makes it an essential part of our daily life. As we aim to reduce the environmental impact of human activity, a historical energy transition is under way. This transition raises several major challenges for electric power systems. We begin with an overview of the general trends of change in power systems, followed by examples of real-world success of mathematical optimization techniques in practice. We then introduce the unit commitment problem and how to obtain commitment decisions that are robust in the context of large-scale penetration of renewables. This is followed by an aggregator-based optimization model to support the participation of so-called prosumers in the electricity markets and their potential to contribute flexibility to the power system. Next we consider several of the recent research developments concerning charging infrastructure for electric vehicles. We conclude with a summary of important future research opportunities for the optimization community in electric energy systems.

SB25

Building A Level 4 A411

Next-Gen Facility Logistics

Invited Session

TSL: Facility Logistics

Chair: Daniela Granados Rivera, Auburn University, Auburn, 36832, United States

1 - Order Picking in Large-Scale Robotic Mobile Fulfillment Systems

Felipe Cordera, Massachusetts Institute of Technology, Cambridge, MA, United States, Alexandre Jacquillat, Cynthia Barnhart, Alexandria Schmid

Robotic process automation is transforming operations across various industries, with e-commerce warehousing increasingly reliant on Robotic Mobile Fulfillment Systems (RMFS) to enable efficient order picking. This paper addresses the challenges of jointly assigning multi-item orders, and the necessary pods, to multiple workstations in large-scale warehouses, through the Task Assignment Problem (TAP). We provide theoretical insights demonstrating that synergy opportunities in order picking expand as system scale increases, underscoring the need to develop scalable solution methods for TAP. To this end, we propose valid inequalities to strengthen its formulation, and we also introduce a family of simple, yet effective, variable-fixing heuristics to tackle large-sized instances. These heuristics, grounded in nearest-neighbors and greedy assignment principles, effectively reduce the dimension of the problem while retaining high quality solutions. Extensive experimentation demonstrates the superiority of our approach over existing benchmarks, further validated through a route planning module that evaluates downstream operational efficiency, specifically focusing on pod traffic bottlenecks. Finally, we introduce a novel optimization framework that integrates downstream operational effects into TAP, enabling more informed decision-making. Our results provide practical insights into challenging operating conditions and showcase the ability of our methods to capitalize on increased synergy opportunities in larger systems.

2 - Leveraging Large Language Models for Intelligent Supply Chain Decision Making

Yongzhi Qi, JD.com, Beijing, China, People's Republic of, Jiaheng Yin, Jianshen Zhang, Wei Qi

We present a novel framework that integrates large language models (LLMs) into intelligent supply chain decision-making, addressing the growing need for responsive, interpretable, and scalable planning systems. Our method has been developed and deployed in collaboration with JD.com, a leading e-commerce platform with highly complex supply chain operations. First, we introduce an LLM-based planning agent capable of autonomously decomposing business goals into executable subtasks across various planning horizons. It coordinates forecasting, inventory control, and replenishment operations by interfacing with external tools and validating output correctness, forming an adaptable and modular planning workflow. Second, we employ LLMs as auditors of traditional decision models. Recognizing that classical optimization methods often rely on unrealistic assumptions, we design prompt-based mechanisms where the LLM critiques decision outputs, surfaces modeling flaws, and suggests business-aligned refinements. This introspective capability improves the practical utility of models in uncertain, data-driven environments. Our experimental results and deployment experience at JD.com show that this hybrid decision-making framework enhances both planning quality and interpretability. It significantly reduces manual effort, improves solution feasibility, and bridges the gap between model-driven and context-aware decision processes. This work demonstrates the practical value of LLMs in enterprise-scale applications beyond text generation. It contributes a robust and extensible paradigm for AI-powered operations, offering new directions for applied research and intelligent system design in the supply chain domain.

3 - On-Demand Service Zone Design for Energy-Constrained Spatial Queueing Systems: Applications to Drones and Mobile Chargers

Peng Lin, Tsinghua University, Beijing, China, People's Republic of, Cheng Hua, Wei Qi, Kai Wang

This research proposes an energy-integrated spatial queueing system that extends the classical hypercube framework with stochastic energy tracking mechanisms. To address energy-constrained service operations, we formulate a semi-Markov process capturing each server's joint energy-position dynamics. Through an iterative approximation scheme with provable fixed-point properties, we numerically derive steady-state service satisfaction rates and operational metrics including system profitability. The implementation framework introduces a service zone design optimization model that simultaneously determines service region partitioning and facility placement under service radius limitations. By constructing the P-median-like problem as an upper bound for our system, we devise an improved branch-and-price algorithm to address the inherent combinatorial complexity. Numerical experiments demonstrate the model's capability in modeling spatial queueing systems with multiple energy-constrained servers operating concurrently, exhibiting resilient performance under server shortage. Specific implementations in multi-task drone operations and mobile charging pile deployment validate its effectiveness in capturing spatial-energy service interactions. The methodology offers systematic insights for designing resilient service networks with adaptive resource pooling capabilities, providing operational guidance for on-demand systems requiring concurrent energy and spatial management.

4 - Mobile Additive Manufacturing Factories with Drone Delivery to Supply Spare Parts to Offshore Platforms

Daniela Granados Rivera, Auburn University, Auburn, AL, United States, Daniel Silva, Alice Smith, Fabio Sgarbossa, Nils Knofius

This research explores a real-world mobile additive manufacturing (MAM) factory problem using uncrewed autonomous vehicles (i.e., drones), from both analytical and practical perspectives. We partner with FIELDMADE, a Norwegian company that provides mobile factories inside shipping containers equipped with additive manufacturing technology. We extend the Mobile Facility Location Problem (MFLP) to study spare parts supply to offshore platforms using MAM factories and delivery drones, enabling a flexible and adaptive supply chain. Our model assumes that MAM factories are located onshore and that delivery options comprise drones, trucks, and ships. The existing literature on MFLP lacks integration of delivery decisions and often limits facility relocation to direct customer locations, which is infeasible for offshore settings. Moreover, the few studies combining MFLP with distribution do not consider multi-modal transportation. Given the constraints of offshore environments, drones are a promising solution for spare parts delivery. The problem incorporates multiple decisions—relocation, production, distribution, and routing—that are rarely addressed together because of the complexity. To address this, we propose a two-stage matheuristic: the first stage focuses on relocation and production, while the second stage addresses distribution, including drone routing and truck/ship deliveries. The performance of the matheuristic proves to be within of 7% of optimal, with computational time up to 20 times faster than the complete mathematical model. Results show that relocation and drone use significantly reduce shipping times and help meet due dates. The proposed approach offers a practical and efficient solution for this novel offshore supply chain.

SB26

Building A Level 4 A412

Data-Driven Insights for Public Sector Operations

Invited Session

Public Sector OR

Chair: Ramin Sepehrirad, Washington State University, PULLMAN, WA, United States

Co-Chair: Chunjie Zhao, Clark University, 950 MAIN STREET, Worcester, MA, 01610-1477, United States

1 - Why is EV Charging so Unreliable? The Role of Competition in Supplementary Services

Yifan Liu, Georgia Institute of Technology, Atlanta, GA, United States

As companies increasingly offer sustainability-oriented supplementary services such as trade-in remanufacturing or hotel electric vehicle charging, understanding the circumstances under which firms can improve their service quality for these offerings is crucial for effective sustainable operations and policy decisions. While competition is widely known to drive improvements in core service quality, its effects on supplementary service quality remain unclear. In this study, we examine how competition influences the reliability of U.S. electric vehicle charging infrastructure, a key supplementary offering for businesses such as hotels, retailers, shopping centers, and parking facilities. Using a U.S. dataset of 66,228 EV charging stations spanning 2011 to 2024, we apply machine learning methods to measure service reliability in EV charging operations. While increased competition within the same core business is found to enhance charging reliability, surprisingly, greater competition from businesses in different core areas is associated with a decline in charging station reliability. This is critical because the vast majority of nearby providers are competing in different core businesses. We project that an average annual increase in competition is associated with a decrease in service reliability scores ranging from 11% to 23% across the most common establishment types over a 5-year period. We also find that the effects of competition attenuate within 10 miles and are not

statistically different from zero in rural or disadvantaged areas. We investigate whether relying on competition, or introducing targeted policies, can address the large-scale service reliability issues reported in the U.S. and globally in public EV charging infrastructure.

2 - Physician Networks in Accountable Care Organizations

Dimitrios Andritsos, HEC Paris, Paris, France, Claire Senot

Effectively coordinating a patient's health care services is critical for the delivery of high-quality and efficient care. Accountable Care Organizations (ACOs) are a prominent model for incentivizing more coordinated care delivery. Nevertheless, no consistent approach has emerged in terms of how different ACOs organize and coordinate care between participating providers. In aggregate, the evidence to date suggests that ACOs' success in achieving their stated objectives of lowering costs and improving quality has been modest. However, this hides the fact that there exists a wide disparity of performance across ACOs. Thus, they provide an excellent empirical setting for the study of care coordination. In this work, we exploit this variability to understand how the nature of care coordination across ACO-centered physician referral networks is associated with performance across both cost and quality dimensions. In our analysis, we put a particular emphasis on coordination between primary care providers and specialists. To characterize communication and care coordination among physicians we examine their patient-sharing relationships via network analysis.

SB27

Building B Level 2 B201

Roundoff-Error-Free Algorithms in Optimization

Invited Session

Computing Society

Chair: Christopher Lourenco, US Naval Academy, Annapolis, MD, United States

1 - An Early Termination Criteria for Dixon's Method

Erick Moreno-Centeno, Texas A&M University, College Station, TX, United States

Dixon's p-adic lifting algorithm uses modular arithmetic to find exact solutions to systems of linear equations. Although this method is state-of-the-art for solving dense systems exactly, it often performs significantly more iterations than necessary to compute the solution. In order to avoid some of these unnecessary computations, a new early termination criteria is presented which is based on Cabay's criteria for congruence technique algorithms. This stopping procedure is predicated on a new version of rational reconstruction which partially reconstructs the modular image found at each p-adic lifting iteration. The complete algorithm maintains the same worst-case complexity as Dixon's original method but is more efficient in practice, as it performs only a small number of iterations beyond those which are necessary to compute the solution.

2 - Sparse Integer-Preserving Gram-Schmidt

Lorena Mejia Domenzain, SimpleRose, St Louis, MO, United States

QR factorization is widely used in mathematics, computer science, and operations research. The QR factorization of matrix A is given by $A=QR$, where Q is an orthogonal matrix, and R is an upper triangular matrix. Typically, QR factorizations are implemented using floating-point precision, which, though appropriate for the majority of applications, have been shown to lead to incorrect rank computation, loss of orthogonality, or incorrect solutions for highly ill-conditioned input matrices. The Round-off error free QR (REF QR) factorization and integer-preserving Gram-Schmidt (IPGS) orthogonalization provide a framework to compute exact QR factorizations of any dense rational input matrix, using integer arithmetic. The REF QR factorization factors matrix A as $A=QDR$, where Q is a left orthogonal matrix, R is an upper triangular matrix, and D is a diagonal matrix. This talk presents an efficient sparse IPGS-REF QR algorithm, that can be used to factorize matrices and determine the rank of rank-deficient matrices. This computation of rank is strongly grounded in exact linear algebra and does not rely on numeric or probabilistic approximations (as other methods for computing the rank of sparse matrices do).

3 - From Pits to Projects: Providing Good Options to Managers Through Computation of K-Best Closures

Andrew Plucker, Texas A&M University, College Station, TX, United States, Erick Moreno-Centeno

Rupert Resources discovered the Ikkari Mineral Reserve in Finland in 2020; their pre-feasibility study highlights a 20-year plan to mine 52 milled tons of earth for an expected gross revenue of \$7.2 billion. The first 10 years of the operation will include open-pit mining. In open-pit mining, the mine design involves selecting blocks of ore to extract while respecting precedence constraints derived from the stability of the pit's slope. When modeled, these constraints form a directed acyclic graph (DAG) where the best economic value corresponds to a maximum weight closure. While the maximum weight closure problem is solvable via a reduction to the minimum-cut problem, many managerial applications demand not just a single optimal solution but a *portfolio* of high-quality alternatives. Generating the k-best closure sets, ordered by total vertex weight, enables decision makers to explore trade-offs, evaluate solution robustness, and flexibly respond to input parameter changes. Despite the practical importance of such ranked alternatives, literature has largely focused on identifying a single optimal closure, with comparatively few tools available for computing and analyzing the k-best solutions. We introduce an algorithmic framework for computing and ranking the k-best closures in a DAG, with direct applicability to managerial problems where solution diversity and interpretability are as important as optimality. Our approach builds upon advances in combinatorial optimization, enhancing them with problem-specific pruning techniques. We demonstrate the utility of our method through instances of open-pit mining and daily-pricing project scheduling, highlighting different solutions that could significantly impact decisions within those respective domains.

4 - Expediting Exact Linear Programming Solvers via Integer Preserving Factorization

Christopher Lourenco, US Naval Academy, Annapolis, MD, United States, Ram Krishnamoorthy

Linear programming (LP) is foundational to operations research and LPs are widely solved using state-of-the-art commercial solvers. These solvers provide accurate solutions for the vast majority of instances; however, for approximately 3-5% of real-world problems, they may provide incorrect solutions. To remedy this issue, exact LP solvers exist, the forefront of which is the solver SoPlex. SoPlex guarantees exactness by performing rational basis validation; that is rational-arithmetic LU factorizations of LP bases. These rational factorizations are the bottleneck of SoPlex, in some cases occupying over 90% of the run time. We replace these rational factorizations with the SPEX factorization framework, an exact integral approach for LU factorization which is asymptotically more efficient than rational LU. On a test set of real-world LPs, our new approach utilizing SPEX for basis validation is a factor of 3.33 and 1.60 times faster than baseline SoPlex in terms of average and geometric mean, respectively.

5 - Stress Testing the Numerical Stability of Lp and Mip Solvers

Joshua Grassel, North Carolina State University, Raleigh, NC, United States, Adolfo Escobedo

Modern solvers for linear programming (LP) and mixed-integer programming (MIP) are indispensable tools in optimization, yet they can yield incorrect results due to roundoff errors from floating-point arithmetic. These errors, such as reporting suboptimal solutions as optimal or misidentifying problems as infeasible, cast doubt on solver reliability. Their impact is particularly acute in domains requiring high numerical precision, such as biochemical modeling, astrophysics simulations, compiler optimization, and mathematical proofs.

To mitigate these issues, various strategies have been developed, including presolve scaling, algorithm tuning, and exact arithmetic subroutines. However, current evaluation practices rely heavily on standard problem libraries, which may not expose numerical weaknesses in a consistent or generalizable way.

This research proposes a novel framework for generating synthetic LP instances that systematically stress test solvers' numerical robustness. Inspired by worst-case complexity constructions like the Klee-Minty cube, the framework allows control over key parameters, such as variable count and coefficient precision, enabling exploration of solver behavior under varying numerical stress. Each generated problem is guaranteed to have an exactly representable optimal solution in the floating-point standard used, ensuring that any observed errors originate from solver internals rather than input and output encoding.

This approach provides a systematic method for testing solver reliability and benchmarking numerical stability. The presentation showcases results from applying this framework to a range of commercial and open-source solvers, highlighting both their strengths and limitations when facing numerically challenging LP instances.

SB28

Building B Level 2 B202

Advances in Healthcare

Invited Session

Health Applications Society

Chair: Yvonne Huijun Zhu, Pennsylvania State University, State College, PA, United States

1 - Harnessing Technological Innovation Through Operational Customization: Virtual Triage Adoption in Acute Care

Jiatao Ding, Nanyang Technological University, Singapore, Singapore, Michael Freeman, Sameer Hasija

Effective technology management is crucial for organizations seeking to leverage technological innovation to improve decision-making and operational performance. While predictive technologies powered by advanced algorithms promise enhanced accuracy and efficiency, real-world evidence suggests that their benefits cannot be realized without proper operational customization – adapting tools to specific organizational contexts. This interplay between technological innovation and operational customization is particularly crucial in acute care, where virtual triage tools have been increasingly adopted to assist patients with self-triage and make informed decisions about the appropriate care options, i.e., primary care or emergency care. Despite their potential as scalable and cost-effective solutions, the impact of virtual triage on patient behavior and system performance remains underexplored. To address this gap, we develop a modeling framework to evaluate the effects of virtual triage adoption, focusing on the nuanced interaction between technological advancement and operational customization. Our analysis reveals a fundamental trade-off between the informativeness and volume of virtual triage recommendations, which significantly influences patient compliance. Additionally, we show that improvements in predictive performance of virtual triage may lead to adverse outcomes. Instead, tailoring virtual triage tools to align with system-specific characteristics offers a pathway to realizing their intended benefits. These findings emphasize the necessity of coupling technological advancements with operational customization to unlock the full potential of virtual triage. More broadly, our work underscores the importance of context-sensitive technology management, calling for regulatory oversight and collaborative efforts between technology providers and operational stakeholders to ensure effective implementations.

2 - Waiting Under Pressure: The Causal Impact of ED Wait Times on Diagnostic Decision-Making

Danqi Luo, UC San Diego, San Diego, CA, United States, Luxi Shen

Emergency department (ED) overcrowding and prolonged wait times pose critical challenges for healthcare systems, yet the direct effect of these waits on clinical decision-making remains unclear. This study combines a natural experiment with a controlled online experiment to investigate whether extended wait times causally influence physicians' diagnostic ordering. Drawing on detailed ED data, we exploit exogenous variation in wait times—conditioned on triage priority and ED congestion—to examine two assumptions: (1) that wait times are independent of underlying patient complexity, and (2) that wait times vary randomly within each acuity level. Empirical tests confirm the validity of these assumptions, showing no significant correlation between wait times and predicted patient complexity, and minimal demographic-based sorting within acuity strata. In a Two-Stage Least Squares (2SLS) analysis, leveraging a ransomware-induced disruption as an instrumental variable, we find that longer waits substantially increase the number of diagnostic tests ordered.

To further validate these findings, we conducted an online experiment on Prolific with practicing clinicians, randomly manipulating only the reported wait time before evaluation. Consistent with the natural experiment results, physicians who perceived a longer patient wait were more likely to order additional tests. These findings highlight the importance of managing ED wait times to optimize diagnostic decisions and improve resource allocation.

3 - What Can Online Personal Statements Tell us? Insights About Physicians's Personality Traits and Clinical Performance

Susan Lu, University of Toronto, West Lafayette, IN, United States, Jianing Ding, Wilson Lin, Karthik Kannan

We use NLP to extract personality traits from physicians' statements and examine how their personalities affect clinical performance.

4 - Explore Then Confirm: Investment Portfolios for New Drug Therapies

Zaile Li, INSEAD, Fontainebleau, France, Stephen Chick, Sam Deams, Shane Henderson

New medical technologies must pass several risky hurdles, such as multiple phases of clinical trials, before market access and reimbursement. A portfolio of technologies pools these risks, reducing the collective financial risk of such development while also improving the chances of identifying a successful technology. We propose a stylized model of a portfolio of technologies, each of which must pass two phases of clinical trials before market access is possible. Using ideas from Bayesian sequential optimization, we study the value of running response-adaptive clinical trials to flexibly allocate resources across technologies in a portfolio. We suggest heuristics for the response-adaptive policy and find evidence for their value relative to non-adaptive policies.

SB29

Building B Level 2 B203

Sector Coupling for the Energy Transition

Invited Session

ENRE: Energy

Chair: Dharik Mallapragada, New York University, Brooklyn, NY, 11201, United States

1 - Industrial Heat Decarbonization Using Gas-Electric Sector-Coupled Models

Dimitri Saad, Stanford University, Stanford, CA, United States, Jacqueline Dowling, Mo Sodwatana, Steve Davis, Inês Azevedo, Adam Brandt

Carbon emissions from industrial heat sources are considered hard-to-abate emissions. Industrial processes generally utilize heat that is generated by combusting a fuel, such as natural gas. Moreover, reducing industrial heating emissions depends on the required process temperatures, which are specific to each industrial sector. Industrial heat electrification can provide net-zero heat at various temperature ranges, provided zero-carbon electricity is utilized. Thus, electrifying industrial heating could require significant expansion in the generation capacity of the electric power grid. Our work examines the role of thermal batteries - a flexible system that converts electricity to heat to store thermal energy - in reducing the process heat emissions of industrial facilities in California. We use BRIDGES: a sector-coupled gas-electricity capacity expansion model, which minimizes system costs over multiple investment years (2025-2045), subject to decreasing carbon emission targets that reach net-zero emissions in 2045. We model heat demands across 39 industrial manufacturing sectors in California, which add up to an equivalent baseload heat demand of 8.1 GW. Our results show that, without the option for carbon capture & storage (CCS) to abate industrial heat emissions, thermal batteries supply the majority of the industrial heat demand below 1000 C. The installation of thermal batteries leads to more built solar PV capacity to charge the batteries. Furthermore, in scenarios with CCS as an option for industrial emission abatement, natural gas usage for industrial heating continues till 2045. While thermal batteries are also built in this scenario, they serve a more substantial role when fuel costs are notably higher.

2 - U.S. Energy Transition Pathways: Insights from US-Regen Scenario Modeling

Aranya Venkatesh, EPRI, Palo Alto, CA, United States, John Bistline

The trajectory of U.S. energy sector emissions over the coming decades will be shaped by multiple dynamic factors, including uncertainties in recent climate legislation, shifts in global fuel trade driven by the energy transition, and the emergence of new technology pathways. In this study, we employ EPRI's US-REGEN—a detailed

energy-economy optimization model that captures the operations and interactions across the electric, fuels, and end-use sectors—to develop scenarios that explore how the U.S. energy system may evolve in response to these drivers. For each scenario, we analyze impacts on the electric sector, focusing on capacity and generation. We also examine key outcomes related to fuel supply, such as the regional distribution and cost of low-carbon fuel production, as constrained by infrastructure and resource availability. Additionally, we assess how end-use demands in the buildings, transportation, and industrial sectors respond to these evolving conditions. Finally, we highlight implications for energy-system CO₂ emissions and system costs.

3 - State-Dependent Non-Linearities in Large-Scale Multi-Sector Energy System Optimization for Capacity Planning

yu weng, Massachusetts Institute of Technology, Cambridge, MA, United States, Ruairidh Macdonald

Multi-sector energy systems; which integrate sectors such as electricity, hydrogen, and transportation; offer a promising pathway to efficiently meet rising energy demands while meeting reliability and decarbonization goals. Capacity expansion planning for such systems are large scale multi-sector co-optimizations of investment and operational decisions across a large number of technologies and resources. As a result, such models can typically only be solved with low temporal or spatial resolution, or simplified technology models - particularly when applied at a national level. To address this, existing methods often rely on clustering and decomposition techniques, along with linearized constraints. While these approaches reduce computational burden, they compromise modeling accuracy, making it difficult to capture the influence of nonlinear behaviors on investment decisions. This work presents an approach to effectively handle non-linear and state-dependent non-linear constraints in planning problems. State-dependent nonlinear constraints is a novel class of constraints that describe how the operational characteristics of technologies/devices change in response to varying levels of investment. These constraints capture important nonlinear properties of technologies that are often overlooked in conventional models. To solve such constraints in capacity planning problem, this work presents a novel envelope-based method for incorporating them within a Benders decomposition framework. The proposed approach approximates state-dependent nonlinear relationships using multiple linear constraints, preserving key nonlinear features without incurring large computational costs. Consequently, it enables a more accurate examination of how the unique properties of emerging technologies and energy carriers could affect future energy investment landscapes.

4 - Modeling Multisectoral Interactions in the Temoa Canadian Open Energy Model

Daniel Posen, University of Toronto, Toronto, ON, Canada, David Elder, Sylvia Sleep, Sean McCoy, Juan Moreno-Cruz, Heather L. MacLean, Sara Hastings-Simon, Andrew Leach, Cameron Wade

Energy system models provide a critical foundation for assessing technology pathways for a low-carbon future. Despite substantial progress in advancing sector-coupling in previously electricity-focused models, many continue to struggle to capturing the sector-specific constraints and multisectoral interactions – particularly in relation to the physical realities such as fuel supply chains, critical mineral and bulk material flows or infrastructure requirements of the energy transition.

This talk will present an overview of ongoing activities to construct and advance a novel energy system optimization model for Canada – the Canadian Open Energy Model (CANOE). Built on the TEMOA linear-programming framework, CANOE is a multi-sector capacity expansion model. It is being constructed in a highly modular format that allows for flexible choice in technological, spatiotemporal and sectoral resolution. The development team is a multi-university consortium with primary background in the tools of industrial ecology (life-cycle assessment, techno-economic assessment, material flow analysis and energy transitions) – enabling us to bring an uncommon perspective to the world of energy modeling. We have developed capabilities to track physical materials (e.g., critical minerals for battery cathodes and electrolyzers) through the energy system; assess the role of embodied carbon in the energy transition; analyze constraints imposed by the available fuel mix from petroleum refineries; incorporate empirical growth-rate constraints on new technologies; and embed results from detailed sectoral models such as transport sector traffic flows and infrastructure needs.

Although the current model remains a coarse representation of the real world, we look forward to discussing our ongoing and future efforts

5 - Domestic Energy System, Macroeconomic, and Distributional Impacts of Increased US Natural Gas Exports

Maxwell Brown, Colorado School of Mines, Golden, CO, United States

This study uses the linked USREP-ReEDS modeling framework to explore the domestic energy system, macroeconomic, and distributional impacts of expanded US natural gas exports. The integration of the computable general equilibrium capabilities of USREP with the detailed electric sector representation of ReEDS helps to capture the relationships between energy infrastructure, economic activity, and household welfare. The analysis explores scenarios of increased natural gas exports, driven by both/either demand and/or supply adjustments, under varying domestic policy conditions. Metrics of focus are energy prices, generation mix, emissions, and regional economic outcomes. Attention is given to the distributional effects across income groups and US regions, highlighting profit and economic welfare changes associated with shifts in energy costs to industrial and residential consumers, respectively. The open-sourcing of the linked model as well as other advancements in sectoral couplings (mostly focused on materials) are also briefly covered.

SB30

Building B Level 2 B204

Understanding Electricity Markets through Enhanced Optimization

Invited Session

ENRE: Electricity

Chair: Alexandra Newman, Colorado School of Mines, Golden, CO, United States

1 - Optimizing Power Grid Expansion Planning to Promote Energy Transition

Yan-Shen Yang, Tianjin University, Tianjin, China, People's Republic of, Bai-Chen Xie

Interregional transmission lines represented by Ultra-High Voltage (UHV) are used to effectively utilize variable renewable energy (VRE) in resource-rich areas and accelerate the low-carbon transformation of the power system. However, considering its high construction cost, whether the transmission line expansion is cost-effective under high VRE penetration remains unknown. Besides, large amounts of power transmission may affect the locally distributed energy development causing the waste of renewable energy resources. Finding an optimal transmission expansion strategy has become a key issue that needs to be solved urgently. To address the above issue, we develop an hourly scale model to simulate power system operations and expansions under different VRE penetrations by 2050. The results show that: (1) Transmission line expansion will decrease the cost of power system under high VRE penetration. The expansion of transmission lines significantly reduces the installed generating capacity of the power system as well as the amount of VRE curtailment. (2) The inter-regional transmission lines have significantly promoted the construction of centralized wind and solar power in the provinces of the “three North” regions, especially in Inner Mongolia and Xinjiang. The distributed PV has developed rapidly in load-concentrated areas with no transmission line expansion. (3) The proposed optimal transmission line expansion can better balance the relationship between interregional power transmission and locally distributed consumption. It achieves cost-effectiveness while avoiding resource wastage. Our study contributes to pursue a smarter decision in the power sector for a sustainable development world.

2 - Grid’S Large Carbon-Abatement Potential Hindered by Interregional Power Market

Bai-Chen Xie, Tianjin University, Tianjin, China, People's Republic of, Yan-Shen Yang

Power transmission lines play an important role in integrating variable renewable energy (VRE) and realizing the low-carbon transformation of power sector. With the low-carbon transformation of power sector, making better use of UHV transmission lines for integrating VRE to decarbonize power sector has become an important issue to be solved urgently. In this study, we have proposed a high-resolution nonlinear programming model for power market that includes independent system operators (ISO),

producers, and consumers, this study finds that the existing transmission lines make no full use of their capacity to improve the utilization of renewable energy. Therefore, we propose two regulatory policies, volume-based and percentage-based, to improve renewable energy power transmission. The results show that (1) the regulatory policy can effectively promote interprovincial VRE transmission, achieving the clean and efficient utilization of power transmission lines. (2) The regulatory policy can optimize the layout of newly installed VRE capacity, as well as significantly promote low-carbon transformation of power sector. (3) The volume-based policy is more effective in promoting carbon emission reduction in the power system than the percentage-based policy. Valuable implications for policymakers are concluded in the end. Our study contributes to the smarter decision in power sector for sustainable development worldwide.

3 - Value-Oriented Forecast Combinations for Unit Commitment

Mehrnoush Ghazanfariharandi, Rutgers University, Piscataway, NJ, United States, Robert Mieth

Value-oriented forecasts for two-stage power system operational problems have been demonstrated to reduce cost, but prove to be computationally challenging for large-scale systems because the underlying optimization problem must be internalized into the forecast model training.

Therefore, existing approaches typically scale poorly in the usable training data or require relaxations of the underlying optimization.

In this talk, we present a method for value-oriented forecast combinations using progressive hedging, which unlocks high-fidelity, at-scale models and large-scale datasets in training.

We also derive a direct one-shot training model for reference and study how different modifications of the training model impact the solution quality.

Our method reduces operation cost by 1.8% on average and trains forecast combinations for a 2736-bus test system with one year of data within 20 hours.

SB31

Building B Level 2 B205

Decarbonization Strategies and Energy Transition

Contributed Session

Contributed

Chair: Ziting Huang, Johns Hopkins University, Baltimore, MD, United States

1 - How to Operate Efficiently and Achieve Maximum Carbon Reduction Before Coal-Fired Power Plants Disappear? a Predict-Then-Optimize Approach

Huiting Hu, University of Science and Technology of China, Hefei, China, People's Republic of

Problem definition: In order to achieve the strategic goal of carbon peak and carbon neutrality in China as soon as possible, it is crucial for coal-fired power plants to effectively utilize the input coals available to improve their fuel costs and reduce carbon emissions. This paper studies the coal blending scheme of coal-fired power plant by mixing different types and corresponding quantities of coal, a coal blending scheme can be obtained under the condition of meeting the demand for power generation. **Methodology/Results:** We developed a data-driven optimal decision framework for coal blending that combines power demand forecasting with coal blending optimization. For this end, we first predict power demand by utilizing the approaches combined long short-term memory (LSTM) with machine learning on multi-modal data. Subsequently, we develop two coal blending optimization formulations, including stochastic optimization formulation and distributionally robust optimization (DRO-W) formulation. To make the formulation solvable, we perform a solvable equivalent reformulation on the proposed DRO-W formulation. By applying our data-driven approach in the numerical experiment based on the historical data of a coal-fired power plant (CPP), which performing significantly better in the worst-case than the stochastic formulation. Furthermore, the numerical results indicate that our data-driven coal blending approach can achieve approximately a 5% cost reduction for the firm. **Managerial Implications:** Interestingly, our findings suggest that the DRO-W approach under carbon trading policy can effectively reduce carbon emission, but the change of carbon price has a non-monotonic effect on the emission reduction.

2 - Structural Features and Formation Mechanisms of Global Co2 Transfer Network Driven by ICT Multinational Investment

Xiaoping Zhang, Tianjin University, Tianjin, China, People's Republic of, Tao Zhao, Rong Yuan, Liang Dong

The rapid expansion of ICT-related multinational enterprises (IMNEs) has introduced significant challenges in emissions mitigation. These IMNEs often manage complex global supply chains and outsource production through foreign direct investment (FDI) while transferring CO₂ emissions to other countries. This paper uses a multi-regional input-output model, social network analysis, and an exponential random graph model to examine the structural characteristics and formation mechanisms of the global CO₂ transfer network driven by IMNEs through FDI (GCNI) from 2000 to 2020. The results reveal that network density increased from 0.353 in 2000 to 0.471 in 2020, reflecting more intensive global CO₂ transfer. The average path length decreased from 1.681 in 2000 to 1.531 in 2020, indicating a shorter distance for embodied CO₂ transfers. The GCNI experienced dynamic cluster adjustments, leading to two distinct communities: the Asia-Africa community and the Europe-America community. China, Chinese Hong Kong, the United States, Japan, South Korea, Germany, and Chinese Taiwan occupied central positions within the GCNI. Larger and more open markets tend to attract CO₂ inflows. CO₂ transfer destinations have shifted from resource-rich nations to those with lower labor costs, and ultimately to regions with high-skilled labor. Countries with strict environmental regulations and high corporate tax rates exhibit a trend of CO₂ outflows. These findings highlight the need to strengthen international cooperation in addressing cross-border environmental challenges in the digital era, as well as the crucial role of FDI motivations in reshaping the geographical patterns of emissions transfers driven by ICT investment.

3 - The Effect of Carbon Prices on Economic Growth and Decarbonization Efforts

Benoit Chevalier-Roignant, emlyon business school, Lyon, France, Alain Bensoussan, Oluwadamilola Oyekan, Viswanath Ramakrishna

The introduction of carbon emission allowances was intended to affect firms' investment decisions (with a possible adverse effect on "economic growth") and allow the emergence and adoption of greener production technologies by incentivizing decarbonization efforts by firms. Empirical studies have stressed the impact of carbon prices' uncertainty on firms' decarbonization efforts. Using a stylized model in continuous time, our model studies when a firm should start operations given that production will imply carbon emissions and, so, the need to acquire certificates. This firm may also choose a production technology with a reduced carbon footprint (e.g., thanks to carbon capture) at the time of investment. In this context, the firm's investment decision is driven by its beliefs about the idiosyncratic profitability and the magnitude of carbon prices. This investment can be reduced to a two-dimensional optimal stopping problem. We characterize the structure of the optimal strategy and design an algorithm to price this real option following a study of a "penalized equation." We also provide comparative statics with respect to key parameters.

4 - Assessing the System Value of Emerging Technologies in A Multi-Sector Energy Model

David Elder, University of Toronto, Toronto, ON, Canada, Juan Moreno-Cruz, Cameron Wade, Heather MacLean, Daniel Posen

In this work, we use an early version of the multi-sectoral TEMOA Canadian Open Energy Model (CANOE) to demonstrate the power of System Value Assessment (SVA) for assessing emerging energy technologies. Rapid decarbonisation requires decisive policy action, which is impeded by risk and uncertainty. A major source of uncertainty in the coming energy transition is the role of emerging technologies like batteries, hydrogen, and direct air carbon capture and storage (DACCS). These technologies feature heavily in plans to reach net zero emissions but there remains deep uncertainty in the projections of their at-scale cost and performance parameters, which are the basis of energy system modelling. This uncertainty is exacerbated by the opacity of large energy systems models, where the interpretation of results can be difficult and often subjective.

Through applied examples exploring the role of grid-scale electricity storage, DACCS, and electrolytic hydrogen production, we demonstrate the use of SVA to assess the value that technologies provide to the energy system across the space of their uncertain parameters – identifying thresholds of price and performance and optimal deployment strategies to achieve and maximise economic viability. We break down the source of system value as a forensic tool to investigate model behaviour, uncovering generalizable insights into the cross-sectoral systemic effects of technology adoption.

5 - Supply Chain and Policy Design for Integrated SAF Production in Brazil: A Bilevel Optimization Approach

Madelynn Watson, University of Notre Dame, Notre Dame, IN, United States, Aline Veronese da Silva, Pedro Gerber Machado, Celma de Oliveira Ribeiro, Claudio Augusto Oller do Nascimento, Alexander Dowling

Sustainable aviation fuel (SAF) is expected to account for 65% of total emissions reductions in the aviation industry by 2050 due to its economic competitiveness and drop-in compatibility with existing aircraft and fuel systems. Brazil's historical government support of biofuels, large ethanol capacity, and existing biomass infrastructure create a unique opportunity for commercial SAF development. Bioethanol from the sugarcane industry in Brazil can be upgraded to SAF via the ASTM-certified pathway alcohol-to-jet (ATJ). Integration of SAF production in the Brazilian sugarcane industry presents complex multiscale optimization problems, including a strong dependence on sustainable energy policy, undeveloped SAF supply chains, and hierarchical interactions between supply chain managers and individual mill owners.

In this work, we propose a bilevel optimization formulation to design a two-stage SAF supply chain in Brazil to meet 50% of the country's international airport jet fuel demand. In this bilevel formulation, we consider the supply chain central planner as the leader. The leader sets the SAF incentive and decides the locations of mills that produce SAF to minimize system costs. Then, the sugarcane mill owners (the followers) determine the production amounts of sugar, ethanol, and SAF to maximize their profit. The optimal supply chain design allows us to determine break-even SAF incentive prices and quantify the impact of ATJ technological advances for successful SAF integration in the Brazilian sugarcane industry.

6 - Plant-Level Techno-Economic Assessment of Steel Sector Decarbonization Pathways in the United States

Robin Zou, Dartmouth College, Hanover, NH, United States, Erin Mayfield

The steel industry accounts for approximately 7% of global greenhouse gas emissions and is widely recognized as a "hard-to-abate" sector. Decarbonizing steel production is essential for meeting climate targets, yet existing analyses often lack plant-level resolution and do not reflect diversity in production technologies, regional material supply chains, and regional energy systems. This gap limits the ability to design tailored, actionable decarbonization strategies.

We develop a bottom-up modeling framework to evaluate the techno-economic pathways across 93 steel-producing plants in the U.S. The model integrates plant-level operational data, technology-specific performance and costs, regional energy system projections, regional material supply chains, and plant-level lifecycle greenhouse gas emissions. Our framework enables comparative assessment of incumbent production routes, such as blast furnace–basic oxygen furnace (BF–BOF), direct reduced iron–electric arc furnace (DRI–EAF), scrap–EAF, and emerging technologies, such as molten oxide electrolysis, carbon capture and storage (CCS), and substitution with low-carbon fuels (e.g., green hydrogen, sustainable biomass).

The framework captures the impact of site-specific variation on cost and emission tradeoffs associated with alternative decarbonization pathways, and examines the role of grid decarbonization, CCS deployment, and fuel switching in shaping feasible transition pathways. Results highlight significant variability in decarbonization potential and cost-effectiveness across plant types and regions.

Our work provides a high-resolution analytical tool for evaluating the economic and environmental implications of decarbonization strategies at the facility level. The framework supports data-driven policy development and strategic planning within the steel sector, helping accelerate industrial decarbonization while addressing operational, technological, and regional constraints.

7 - A Study on the Economic Impacts of Repurposing Plant Sites in Energy Transition

Jeongmin Shin, Kyung Hee University, Yongin-si, Korea, Republic of, Changeun Park, Jungwoo Shin

As coal-fired power plants are phased out globally to support decarbonization, South Korea faces both challenges and opportunities in repurposing the idle land left behind—often near urban centers. This study presents a comprehensive analysis of such repurposing by integrating three dimensions: social preference, economic value, and financial impact on public energy companies. A nationwide choice experiment using a Bayesian mixed logit model reveals that residents prefer tourism-oriented or hybrid redevelopment over simple power plant conversions, driven by factors like carbon reduction, job creation, and proximity. The zonal travel cost method, applied to the Mapo Saebit Culture Forest, estimates significant consumer surplus from nature-based tourism, projecting long-term benefits for nine closure sites. Financial analysis using a Korean-adapted O-score model shows that although closure compensation harms short-term financial stability, long-term recovery is achievable through electricity revenues, carbon savings, and tourism income. The study offers a novel framework combining stated preference data, economic valuation, and financial risk modeling to inform site-specific, regionally sensitive policy. While it delivers valuable insights, it acknowledges limitations, including the exclusion of site-specific constraints and fixed financial assumptions. Overall, the research highlights that repurposing idle energy infrastructure requires a multidimensional strategy that blends physical planning with socioeconomic and financial considerations, offering practical guidance for sustainable energy transition and regional resilience.

8 - Location-Based Power Planning with Aging Plant Sites for a Just Energy Transition: Coal Repurposing Opportunities in South Africa

Ziting Huang, Johns Hopkins University, Baltimore, MD, United States, Ben Hobbs, Debabrata Chattopadhyay

By 2045, 25% of the global operating coal power capacity will reach the end of 50-year design lifetime, presenting a critical crossroads for system operators and plant owners in making future decisions. However, plant sites remain valuable assets, as transmission lines typically have a much longer lifespan than generators. To explore the reutilization potential of plant sites, we propose a holistic model that compares the benefits across six alternatives regarding existing coal power: (1) limit operation, (2) mothball as a backup generator, (3) retrofit with carbon capture and storage, (4) repurpose with renewable energy and battery storage, (5) redevelop for non-power use, and (6) demolish completely. These decarbonizing alternatives often come along with different externalities, such as providing affordable generation, reducing carbon emissions, or increasing unemployment.

We incorporate multiple objectives with capacity expansion model to identify the plant redevelopment decisions while achieving system stability, emissions reduction, and local employment. Furthermore, system operators typically prioritize costs and system reliability, while plant owners are more concerned with profits and local community impacts, leading to discrepancies in objectives during the decision-making process. We also explore the interaction between system and plant owners through two distinct modeling frameworks -- multi-objective optimization and agent-based modeling simulation, to facilitate a comparison between top-down decisions and those incorporating with plant-level preferences. This research is useful in developing effective energy transition policies that balance both system stability and local community benefits.

We adopt this model in South Africa, a system dominated by aging coal, to discuss coal phase-out decisions.

SB32

Building B Level 2 B206

Navigating Deep Climate Uncertainty in Energy Modeling

Invited Session

ENRE: Environment and Sustainability

Chair: Fariba Farajbakhsh Mamaghani, Tulane University, River Ridge, LA, United States

1 - The Impact of Precipitation Forecasts on Small Business Performance

Weilong Wang, Columbia University, New York, NY, United States, Tim Sels, Abhishek Nagaraj, Vaibhav Anand

Weather forecasts are a critical tool for economic decision-making, yet systematic evidence on their real-world value remains limited. This study quantifies the economic impact of short-term precipitation forecasts on small business performance across the United States. Using a natural experiment framework, we exploit variation in forecast accuracy while holding actual weather constant to isolate the causal effect of forecast quality on business outcomes. Specifically, we examine how customer foot traffic and spending at over one million establishments vary when precipitation is predicted accurately versus when forecasts either over- or under-estimate rainfall. Drawing on high-frequency data from SafeGraph and anonymized transaction datasets, we find that forecast errors meaningfully influence small business activity. On days when rainfall is overpredicted, businesses see a measurable decline in visits and spending—even when actual conditions are favorable. Conversely, underpredictions lead to increased business activity, though often at the cost of preparedness. These effects are concentrated on days with moderate to heavy precipitation and vary significantly by industry, in particular, retail, food service, and entertainment. Our results highlight the economic cost of inaccurate forecasts and the value of reliable short-term predictions. Improvements in 1-day precipitation forecasts yield substantial returns in increased consumer activity during inclement weather, suggesting investment in forecast accuracy significantly benefits local economies. By providing causal evidence, this study offers a rigorous framework for evaluating forecasts and underscores the importance of incorporating user impact into forecasting system design and funding decisions.

2 - Facing Common Threats in a Polarized Environment

Cathy DiGennaro, MIT Sloan School of Management , Cambridge, MA, United States, Vicky Chuqiao Yang

When facing a common threat, conventional wisdom suggests that members of a society will be able to set aside their differences and work together towards a collective solution. However, that this axiom is universal has been disproven by the disjointed and highly polarized responses to COVID-19 and the threat of climate change in the United States. Previous research has suggested that conditions are unfavorable for collective response, citing high levels of affective polarization, low information and subsequent overreliance on social learning, and divisive influence of political elites. We synthesize these concepts into a mathematical model of opinion dynamics and evaluate different scenarios under which feedback between affective and ideological polarization may evolve. Under different specifications for social learning, initial polarization (representing, e.g., early influence by political elites), and inherent preferences between groups, the model predicts three behaviors. In a scenario with a high degree of social learning and high initial polarization, a common threat can deepen existing division, even when inherent preferences are equal between two groups. When initial polarization is lower, groups can reach consensus on the common threat, and when reliance on social learning is reduced and inherent preferences are more disparate, the consensus reached on the common threat can reduce polarization on pre-existing issues. This study helps solve the puzzle of why common, catastrophic threats like COVID-19 and climate change are not guaranteed to unite polarized groups and suggests potential high-leverage intervention points to facilitate collective problem-solving.

3 - Electrifying Away Climate Change: Interactions of Heat Pump Adoption, Installation Capacity, and Peak Grid Load in Massachusetts

Alexander Kuptel, MIT Sloan School of Management , Cambridge, MA, United States, John Sterman

The transition to electric heat pumps for residential heating is likely an effective step in addressing climate change. Despite their efficiency and emissions benefits, heat pump adoption remains limited due to high upfront costs and insufficient installer capacity. Moreover, their performance in cold climates raises concerns about grid load during peak winter demand, and subsidizing heat pumps has proven to be expensive. Using a dynamic behavioral model parameterized to Massachusetts, this paper identifies high leverage policies that spur heat pump adoption while limiting their effect on grid load and costs. Our results suggest subsidies for heat pump purchase costs and electricity use only minimally boost adoption due to limited installation capacity. Subsidizing electricity use also disincentivizes home energy retrofits, as homeowners face a lower effective heating cost. Homes under greater subsidies are subsequently less efficient, leading to greater energy use, peak winter load, and heating energy subsidies. On the other hand, a policy that targets operational and financial constraints of heat pump adoption, such as funding more installer training, improves adoption. This policy increases adoption especially in the short run and therefore reduces cumulative emissions significantly compared to the subsidy policy. However, higher short-run adoption also increases peak heating load and subsidies, highlighting a tradeoff of adoption and emissions with peak winter load and costs in the short term. Our results emphasize the importance of installation capacity and the interaction between energy costs and home energy retrofits to accelerate the energy transition in residential heating.

4 - Bilevel Analysis of Cost and Emissions Externalities from Data Center Load Shifting

Aron Brenner, MIT, Cambridge, MA, United States, Rahman Khorramfar, Nathan Engelman Lado, Line Roald, Saurabh Amin

Data centers are emerging as large, flexible electricity consumers capable of shifting computational workloads across locations in response to economic and environmental signals. While this flexibility has potential for emissions reduction, its impact on power system operations depends critically on how such behavior interacts with network constraints and market signals. We develop a bilevel optimization framework in which a data center minimizes a weighted combination of electricity cost and marginal emissions intensity (LME), while the system operator clears economic dispatch under transmission and generation constraints. Focusing on a stylized three-bus power system, we derive closed-form, piecewise-linear expressions for both the data center and system-wide objectives as functions of the data centers' load shift. These expressions capture threshold-driven regime changes due to congestion and renewable saturation. We identify sufficient conditions under which the data center's decentralized decisions align with or diverge from socially optimal behavior and characterize the resulting externalities. Our results reveal how system topology and generator asymmetry affect incentive alignment and provide insight into when marginal price or emissions signals may fail to guide flexible loads toward socially beneficial outcomes. Our results offer a tractable starting point for analyzing decentralized flexibility under carbon-aware incentives and suggest directions for improving coordination between flexible loads and system operations.

SB33

Building B Level 2 B207

Healthcare NLP: Leveraging Language Models for EHR Insights

Invited Session

Health Applications Society

Chair: Muyan Jiang, UC Berkeley, Berkeley, CA, United States

Co-Chair: Anil Aswani, UC Berkeley, Berkeley, CA, United States

1 - A Large Language Model Based Pipeline for Review of Systems Entity Recognition from Clinical Notes

Zhuqi Miao, SUNY New Paltz, New Paltz, NY, United States, Hieu Nghiem, Hemanth Singareddy, Abdulaziz Ahmed, Jivan Lamichhane, Dursun Delen, William Paiva

The Review of Systems (ROS) is a critical component of clinical assessments, helping identify and differentiate symptoms across body systems. While ROS data are often documented as free text in clinical notes, this format poses challenges for analysis and contributes to physician documentation burden. Despite the promise of NLP for parsing such text, most existing research relies on complex, fine-tuned BERT models rather than more user-friendly large language models (LLMs). This study introduces a streamlined, cost-effective pipeline using open-source LLMs for ROS recognition, offering a scalable solution for improving medical documentation workflows.

2 - Uncovering Treatment Non-Adherence Bias in Clinical Machine Learning with LLMs: a Real-World Study on Hypertension Medication

Zhongyuan Liang, UC Berkeley, Berkeley, CA, United States

Machine learning systems trained on electronic health records (EHRs) increasingly guide treatment decisions, but their reliability depends on the critical assumption that patients follow the prescribed treatments recorded in EHRs. Using EHR data from 3,623 hypertension patients, we investigate how treatment non-adherence introduces implicit bias that can fundamentally distort both causal inference and predictive modeling. By extracting patient adherence information from clinical notes using a large language model (LLM), we identify 786 patients (21.7%) with medication non-adherence. We further uncover key demographic and clinical factors associated with non-adherence, as well as patient-reported reasons including side effects and difficulties obtaining refills. Our findings demonstrate that this implicit bias can not only reverse estimated treatment effects, but also degrade model performance by up to 5%, while disproportionately affecting vulnerable populations by exacerbating disparities in decision outcomes and model error rates. This highlights the importance of accounting for treatment non-adherence in developing responsible and equitable clinical machine learning systems.

3 - Scaling Clinical Note Understanding: An LLM-Guided Framework for Aspect-Based Sentiment Analysis

Muyan Jiang, UC Berkeley, Berkeley, CA, United States

We introduce a flexible, LLM-enhanced framework for aspect-based sentiment analysis of clinical narratives. Our approach begins by sampling representative note excerpts across specialties and time windows, then engages a large language model to define and illustrate key clinical aspects—such as empathy, uncertainty, patient-centered explanations, shared decision-making, and confidentiality. Leveraging few-shot prompting, the same LLM rapidly generates labeled examples, which we use to fine-tune a single transformer classifier. At scale, the trained model assigns sentiment scores for each aspect across entire clinical notes, handling long texts through intelligent segmentation. The result is a compact, per-note sentiment profile that reflects multiple dimensions of clinician language. This end-to-end pipeline strikes a balance between human-guided design and automated scalability, making it readily adaptable for large clinical corpora.

SB34

Building B Level 2 B208

Logistics Under Stress

Invited Session

OPT: Global Optimization

Chair: Daniel Bienstock, Columbia University, New York, NY, 10027, United States

1 - A Column-Generation Approach for Resilient Logistics in Adversarial Environments

Gabriel Hake, Colorado School of Mines, Golden, CO, United States, Robert Provine, Justin Kilb, Alexandra Newman, Daniel Bienstock

This study integrates column generation into a multi-period contested logistics framework, allowing Blue forces to dynamically adapt to adversary actions. Traditional logistics models assume predictable movement, yet adversarial conditions introduce uncertainty and disruption beyond conventional forecasting methods. Our approach iteratively refines feasible solutions, leveraging Generative Adversarial Tactics for Engagement to assess adversarial behaviors without presupposing static distributions. By solving an adversary-aware routing and inventory management problem, we generate robust strategies that optimize resupply efforts while responding to emerging threats. Joint work with Bobby Provine, Justin Kilb, Blake Sisson, Alexandra Newman and Daniel Bienstock.

2 - RI for Logistics Under Stressed Conditions

Blake Sisson, Columbia University, New York, NY, United States, Daniel Bienstock, Justin Klib, Alexandra Newman

We consider a setup with sequential decision making where a logistics plan is stressed by a poorly understood adversary. We demonstrate a set of computational tools based on optimization and reinforcement learning, designed to obtain resilient decisions. Joint work with Justin Kilb, Alexandra Newman and Daniel Bienstock.

3 - A MIP-Based Planning Tool for Military Logistics Operations

Soochan Kim, University of Southern California, Los Angeles, CA, United States, Johannes Royset

In this presentation, we propose a mixed integer programming-based decision support tool that optimizes military logistics operations. The mathematical formulation aims to maximize the fulfilled transport requests, each defined by commodity type, quantity, and a priority factor, across an area of operations consisting of several nodes, while minimizing connector and crew usage cost. In addition, the proposed model considers complex real-world constraints such as capacity, range, refueling, and crew assignment. Computational experiments on real-scale instances indicate that the model efficiently produces sustainment plans within a reasonable computation time. We also explore alternative objective function mode and constraint settings to adapt the tool to varying operational situations and discuss implications of the results.

4 - Tactical Route Planning

Jakub Cerny, Columbia University, NYC, NY, United States, Christian Kroer, Garud Iyengar

We build a game-theoretic model for tactical-level route planning in contested environments with a strategic adversary. While basic deterministic logistics plans are already challenging to devise, they tend to be predictable, which can limit their effectiveness. By introducing calculated randomness into the route planning process, we compute immediately deployable plans that are diversified and harder to anticipate, while maintaining operational efficiency through computation times of just a few seconds. Joint work with Christian Kroer.

5 - Solving Large-Scale Military Logistics Planning Problems

Jeff Linderorth, University of Wisconsin-Madison, Madison, WI, United States

Dreamcatcher-Log (dclog) is a mixed-integer linear program that implements a multi-commodity flow model over a time-space network to propose best allocation and scheduling of connector resources to meet the force closure and sustainment requirements in the logistics plan to support operations meeting the commander's intent. In this talk, we outline the dclog mixed-integer programming model and discuss recent computational improvements.

SB35

Building B Level 2 B209

Causal Inference: Data Corruption, Debiasing, and Fairness.

Invited Session

Public Sector OR

Chair: Jimi Kim, University of Texas at Dallas, Richardson, TX, United States

1 - Learning Interpretable Concepts: Unifying Causal Representation Learning and Foundation Models

Bryon Aragam, University of Chicago, Chicago, IL, United States

To build intelligent machine learning systems, there are two broad approaches. One approach is to build inherently interpretable models, as endeavored by the growing field of causal representation learning. The other approach is to build highly-performant foundation models and then invest efforts into understanding how they work. In this work, we relate these two approaches and study how to learn human-interpretable concepts from data. Weaving together ideas from both fields, we formally define a notion of concepts and show that they can be provably recovered from diverse data. Experiments on synthetic data and large language models show the utility of our unified approach.

2 - Prompting Fairness: Causality-Guided Debiasing of LLM Decisions

Zeyu Tang, Stanford University, Stanford, CA, United States

Large language models (LLMs), despite their remarkable capabilities, are susceptible to generating biased and discriminatory responses. As LLMs increasingly influence high-stakes decision-making (e.g., hiring and healthcare), mitigating these biases becomes critical. We propose a causality-guided debiasing framework to tackle social biases, aiming to reduce the objectionable dependence between LLMs' decisions and the social information in the input. Our framework introduces a novel perspective to identify how social information can affect an LLM's decision through different causal pathways. Leveraging these causal insights, we outline principled prompting strategies that regulate these pathways through selection mechanisms. This framework not only unifies existing prompting-based debiasing techniques, but also opens up new directions for reducing bias by encouraging the model to prioritize fact-based reasoning over reliance on biased social cues. Extensive experiments on real-world datasets across multiple domains have demonstrated the effectiveness of our framework in debiasing LLM decisions, even with only black-box access to the model.

3 - Unified Covariate Adjustment for Causal Inference

Yonghan Jung, University of Illinois at Urbana-Champaign (UIUC), Urbana-Champaign, IL, United States

Causal effect identification and estimation are two crucial tasks in causal inference. Although causal effect identification has been theoretically resolved, many existing estimators only address a subset of scenarios, known as the sequential back-door adjustment (SBD) (Pearl and Robins, 1995) or g-formula (Robins, 1986). Recent efforts for developing general-purpose estimators with broader coverage, incorporating the front-door adjustment (FD) (Pearl, 2000) and more, lack scalability due to the high computational cost of summing over high-dimensional variables. In this paper, we introduce a novel approach that achieves broad coverage of causal estimands beyond the SBD, incorporating various sum-product functionals like the FD, while maintaining scalability -- estimated in polynomial time relative to the number of variables and samples. Specifically, we present the class of UCA for which a scalable and doubly robust estimator is developed. In particular, we illustrate the expressiveness of UCA for a wide spectrum of causal estimands (e.g., SBD, FD, and more) in causal inference. We then develop an estimator that exhibits computational efficiency and doubly robustness. The scalability and robustness of the proposed framework are verified through simulations.

4 - Near-Optimal Reinforcement Learning in Dynamic Treatment Regimes

Junzhe Zhang, Syracuse University, Syracuse, NY, United States, Elias Bareinboim

A dynamic treatment regime (DTR) consists of a sequence of decision rules, one per stage of intervention, that dictates how to determine the treatment assignment to patients based on evolving treatments and covariates' history. These regimes are particularly effective for managing chronic disorders and is arguably one of the key aspects towards more personalized decision-making. In this paper, we investigate the online reinforcement learning (RL) problem for selecting optimal DTRs provided that observational data is available. We develop the first adaptive algorithm that achieves near-optimal regret in DTRs in online settings, without any access to historical data. We further derive informative bounds on the system dynamics of the underlying DTR from confounded, observational data. Finally, we combine these results and develop a novel RL algorithm that efficiently learns the optimal DTR while leveraging the abundant, yet imperfect confounded observations.

5 - Optimizing Policy Decisions in Interconnected Populations: a Spatially-Dependent Framework for Dynamic Decision-Making

Jimi Kim, University of Texas at Dallas, Addison, TX, United States, Yunan Wu

Dynamic decision-making is crucial in policy formation. In reality, states or countries typically make these decisions, and their effects are interdependent due to spatial relationships. However, conventional policy decision-making strategies often overlook these spatial effects, assuming no interference among individuals. In this project, we propose a novel framework to measure these interdependent effects across multiple stages. We also propose the methods to estimate the optimal policy regime taking the spatial effects in consideration for a series of time spots, as well as constructing confidence intervals for the coefficients indexing the optimal policy regime based on multiplier Bootstrap methods. We rigorously demonstrate the validity of the proposed estimation and inference methods. We apply the proposed method on a case study on COVID-19 school closing policies. By applying our methodologies to real-world data on state-level interventions and outcomes, we illustrate how this approach can optimize policy decisions. Our findings have substantial implications for enhancing the effectiveness of personalized intervention strategies in interconnected populations. This approach has the potential to revolutionize precision policy-making in spatially dependent contexts while maintaining a balance between public health and economic considerations.

SB36

Building B Level 2 B210

Advances in Learning Dynamic Games

Invited Session

OPT: Nonlinear Optimization

Chair: Apurv Shukla, University of Michigan, Ann Arbor, MI, United States

1 - Capacitated Dynamic Pricing with Asymmetric Information

Apurv Shukla, University of Michigan, Ann Arbor, MI, United States

We consider the problem of capacitated dynamic pricing under information asymmetry. Our main contribution is to study the capacitated dynamic pricing problem under information asymmetry using the notion of sequential equilibria. We study sequential equilibria as such refinements are more credible in dynamic games than Bayes-Nash equilibrium. We establish the existence of this equilibrium for our problem and propose an efficient method to compute a subset of them. We study characteristics of the resulting pricing equilibrium, such as existence, uniqueness, and the effects of competition and information asymmetry. We conclude by examining the efficiency of the proposed equilibrium with respect to two benchmarks: the team problem which preserves information asymmetry and a mechanism design perspective which computes the transfers and utilities by soliciting information apriori. In addition to providing insights into the efficiency of the pricing equilibria, these benchmarks also shed light on the effect of information asymmetry on social welfare in such dynamic pricing problems.

2 - The Lagrangian Method for Solving Constrained Markov Games

Soham Das, Texas A&M University, College Station, TX, United States, Santiago Paternain, Luiz Chamon, Ceyhun Eksin

We introduce the conception of a Lagrangian game to solve general sum constrained Markov games (CMGs). CMGs model multiagent systems in dynamically evolving environments where agents optimize their individual utilities while also adhering to global system-level feasibility constraints. These constraints can model real world limitations, such as resource budgets, safety considerations, etc. CMGs provide the formal underpinning for safe multiagent reinforcement learning (safe-MARL), where agents must learn intrinsically safe policies through repeated interactions in a coupled environment. Applications are ubiquitous, including autonomous systems coordination, energy aware planning, etc. In our proposed algorithmic framework, given a general-sum CMG, agents construct the “dual” Lagrangian game, by augmenting their value functions akin to constructing the Lagrangian in constrained optimization. Agents solve the associated unconstrained Lagrangian game and simulate finite state-action trajectories to evaluate accumulated rewards and costs. The experience thus obtained is utilized to update the multipliers using a gradient step, generating a new Lagrangian game iteration. Our key result consists in showing that the sequence of policies thus generated constitutes a nonstationary approximate Nash for the original CMG. Our results guarantee feasibility for the policy sequence, with controlled accuracy, while handling constraints indirectly.

3 - Sharing with Frictions: Limited Transfers and Costly Inspections

Federico Bobbio, Northwestern University, Chicago, IL, United States, Randall Berry, Michael Honig, Thanh Nguyen, Vijay Subramanian, Rakesh Vohra

We study an optimal mechanism design problem for resource sharing between an incumbent noncommercial user and a commercial entrant. Motivated by applications such as wireless spectrum allocation and access to national resources, where noncommercial and commercial users compete for limited availability, the mechanism is subject to two key constraints. First, direct transfers cannot take place to compensate the incumbent for externalities caused by the activity of the commercial user. Instead, the incumbent reports a disutility for sharing to a regulator; the second constraint is that this disutility can only be verified via a costly inspection. The regulator optimizes total welfare by announcing a probability of shared resource assignment along with a probability of inspection. We show that in the optimal mechanism, both allocation and inspection decisions are binary: The regulator either fully allows or prohibits sharing, and inspections occur only for certain reported values when sharing is denied. We further reformulate the problem as a knapsack model to characterize threshold extremal points.

SB37

Building B Level 2 B211

Complexity Analysis for Deterministic, Stochastic, and Robust Optimization Methods

Invited Session

OPT: Optimization Under Uncertainty

Chair: Hongcheng Liu, University of Florida, Gainesville, FL, United States

1 - A Single-Loop Spider-Type Stochastic Subgradient Method for Expectation-Constrained Nonconvex Nonsmooth Optimization

Yangyang Xu, Rensselaer Polytechnic Institute, Troy, NY, United States, Wei Liu

We consider nonconvex nonsmooth stochastic optimization problems with expectation constraints, for which we build a novel exact penalty model. We first show the relationship between the penalty model and the original problem. Then on solving the penalty problem, we present a single-loop SPIDER-type stochastic subgradient method, which utilizes the subgradients of both the objective and constraint functions, as well as the constraint function value at each iteration. Under certain regularity conditions (weaker than Slater-type constraint qualification or strong feasibility assumed in existing works), we establish an iteration complexity result that is better than the best-known results. Moreover, on solving two fairness-constrained problems, our method is significantly (up to 466 times) faster than the state-of-the-art algorithms, including switching subgradient method and inexact proximal point methods.

2 - Parameter-Free Gradient Sliding Methods

Yuyuan Ouyang, Clemson University, Clemson, SC, United States

The gradient sliding methods was developed for problems that has separable structures: sum of convex smooth and nonsmooth functions, convex smooth objective and constraint functions, etc. The core idea of gradient sliding-type methods is that the complexity of first-order algorithms should be adaptable to different problem structures. While several gradient sliding methods have been proposed in the literature, few of them are parameter-free; most sliding algorithms require the knowledge of Lipschitz constants. In this talk, we show that gradient sliding methods can be designed without any knowledge of Lipschitz constants. We propose novel gradient sliding algorithms for convex functional constrained minimization and monotone variational inequalities with convex functional constraints.

3 - Simple yet Effective Approximations for Optimization Under Uncertainty

Jianqiang Cheng, University of Arizona, Tucson, AZ, United States

In this talk, we present a novel and simple modeling method called harmonizing optimization (HO), which integrates SAA and DRO with moment by adaptively adjusting the weights of data and information based on sample size N . This allows HO to amplify data effects in large samples while emphasizing information in smaller ones. Moreover, we provide practical methods for determining these weights and demonstrate that HO offers finite-sample performance guarantee. As an application, HO can be used to enhance scenario reduction by retaining critical information from reduced scenarios, thereby improving approximation quality and reducing completion time. Numerical results show significant advantages of HO in solution quality compared to other benchmark modeling approaches, and highlight its effectiveness in scenario reduction.

4 - Metric Entropy-Free Sample Complexity Bounds for Sample Average Approximation in Convex Stochastic Programming

Hongcheng Liu, University of Florida, Gainesville, FL, United States

This talk is focused on sample average approximation (SAA) in solving convex or strongly convex stochastic programming (SP) problems. Under some common regularity conditions, we show that SAA's sample complexity can be completely free from any quantification of metric entropy (such as the logarithm of the covering number), leading to a significantly more efficient rate with dimensionality than most existing results. From the newly established complexity bounds, an important revelation is that SAA and the canonical stochastic mirror descent (SMD) method, two mainstream solution approaches to SP, entail almost identical rates of sample efficiency, lifting a current theoretical discrepancy of SAA from SMD by the order of $O(d)$.

SB39

Building B Level 2 B213

Online and Discrete Optimization: Methods and Applications

Invited Session

OPT: Integer and Discrete Optimization

Chair: Alexandre Jacquillat, MIT Sloan School of Management, Cambridge, MA, United States

1 - Online Rack Placement in Large-Scale Data Centers: Online Sampling Optimization and Deployment

Alexandre Jacquillat, MIT Sloan School of Management, Cambridge, MA, United States, Kayla Cummings, Sean Lo, Konstantina Mellou, Ishai Menache, Marco Molinaro

This paper optimizes the configuration of large-scale data centers toward cost-effective, reliable and sustainable cloud supply chains. The problem involves placing incoming racks of servers within a data center to maximize demand coverage given space, power and cooling restrictions. We formulate an online integer optimization model to support rack placement decisions. We propose a tractable online sampling optimization (OSO) approach to multi-stage stochastic optimization, which approximates unknown parameters with a sample path and re-optimizes decisions dynamically. We prove that OSO achieves a strong competitive ratio in canonical online resource allocation problems and sublinear regret in the online batched bin packing problem, and that it can outperform certainty-equivalent benchmarks. Our algorithm has been packaged into a software solution deployed across Microsoft's data centers, contributing an interactive decision-making process at the human-machine interface. Using deployment data, econometric tests suggest a positive and statistically significant impact of our solution on data center performance, leading to a 1--3 percentage point reduction in power stranding. At the scale of cloud computing, these improvements in data center utilization can result in multi-million-dollar annual cost savings and concomitant environmental benefits.

2 - Learning to Cover: Online Learning and Optimization with Irreversible Decisions

Michael Lingzhi Li, Harvard Business School, Boston, MA, United States, Alexandre Jacquillat

We define an online learning and optimization problem with discrete and irreversible decisions contributing toward a coverage target. In each period, a decision-maker selects facilities to open, receives information on the success of each one, and updates a classification model to guide future decisions. The goal is to minimize facility openings under a chance constraint reflecting the coverage target, in an asymptotic regime characterized by a large target number of facilities $m \rightarrow \infty$ but a finite horizon $T \in \mathbb{Z}^+$. We prove that, under statistical conditions, the online classifier converges to the Bayes-optimal classifier at a rate of at best $O(n^{-1/2})$. Thus, we formulate our online learning and optimization problem, with a generalized learning rate $r > 0$ and a residual error $1 - p$. We derive an asymptotically optimal algorithm and an asymptotically tight lower bound. The regret grows in $\Theta(m(1-r)/(1-r^T))$ if $p=1$ (perfect learning) or in $\Theta(m \max\{(1-r)/(1-r^T), 1/2\})$ otherwise; in particular, the regret rate is sub-linear and converges exponentially fast to its infinite-horizon limit. We extend this result to a more complicated facility location setting in a bipartite facility-customer graph with a target on customer coverage. Throughout, constructive proofs identify a policy featuring limited exploration initially and fast exploitation later on once uncertainty gets mitigated. These results uncover the benefits of online learning and optimization through pilot programs prior to full-fledged expansion.

3 - Sequential Testing with Subadditive Cost

Rayen Tan, University of Michigan, Ann Arbor, MI, United States, Blake Harris, Viswanath Nagarajan

In the classic sequential testing problem, we are given a system with several components each of which fails with some independent probability. The goal is to identify whether or not some component has failed. When the test costs are additive, it is well known that a greedy algorithm finds an optimal solution. We consider a much more general setting with subadditive cost functions and provide a $(4q+\gamma)$ -approximation algorithm, assuming a γ -approximate value oracle (that computes the cost of any subset) and a q -approximate ratio oracle (that finds a subset with minimum ratio of cost to failure probability). While the natural greedy algorithm has a poor approximation ratio in the subadditive case, we show that a suitable truncation achieves the above guarantee. Our analysis is based on a connection to the minimum sum set cover problem. As applications, we obtain the first approximation algorithms for sequential testing under various cost-structures: $(5+\epsilon)$ -approximation for tree-based costs, 9.5-approximation for routing costs and $(4+\ln n)$ for machine activation costs. We also show that sequential testing under submodular costs does not admit any poly-logarithmic approximation (assuming the exponential time hypothesis).

4 - New Optimization Approaches for Designing Experiments

Zulal Isler Ardic, Rensselaer Polytechnic Institute, Troy, NY, United States, Diego Moran, Burak Kocuk

Designing experiments is essential for drawing reliable conclusions in research. A key challenge lies in dividing participants into groups that are both balanced and representative of the population. This ensures that comparisons between treatments or controls are fair and meaningful. Traditional methods, such as randomization and re-randomization, are widely used to reduce bias. However, these approaches do not guarantee optimal group balance, particularly in studies with small or diverse samples. Optimization-based methods have emerged as a promising alternative, offering improved performance by directly targeting several balance criteria. In this talk, we present new optimization models that combine different metrics such as difference in moments, Wasserstein distance, variation distance, and a new bin-based approach. Through computational experiments, we evaluate the strengths and limitations of each approach in achieving group similarity and empirically show that our models have a promising performance.

5 - Rounding the Lovasz Theta Function with a Value Function Approximation

Rui Gong, Georgia Institute of Technology, Atlanta, GA, United States, Diego Cifuentes, Alejandro Toriello

The Lovasz theta function is a semidefinite programming (SDP) relaxation for the maximum weighted stable set problem, which is tight for perfect graphs. However, even for perfect graphs, there is no known rounding method guaranteed to extract an optimal stable set from the SDP solution.

In this paper, we develop a novel rounding scheme for the theta function that constructs a value function approximation from the SDP solution

and then constructs a stable set using dynamic programming.

Our method provably recovers an optimal stable set in several sub-classes of perfect graphs, including generalized split graphs, which asymptotically cover almost all perfect graphs.

To the best of our knowledge, this is the only known rounding strategy for the theta function that recovers an optimal stable set for large classes of perfect graphs.

Our rounding scheme relies on simple linear algebra computations;

we only solve one SDP.

In contrast, existing methods for computing an optimal stable set in perfect graphs require solving multiple SDPs.

Computational experiments show that our method produces good solutions even on imperfect graphs.

SB40

Building B Level 2 B214

Healthcare Data Analytics

Invited Session

Service Science

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

1 - Llm-Driven Preprocessing of Ehr and Survey Data for Medical Downstream Applications

Kihyuk Yoon, Georgia Institute of Technology, Atlanta, GA, United States, Lingchao Mao, Jing Li

In this study, we propose a preprocessing method that generates embedding vectors for downstream medical machine learning tasks using large language models (LLMs). This approach addresses common challenges in handling electronic health records (EHRs) and medical survey data, such as high rates of missing values and categorical variables with extensive, sparse categories like medications or comorbidities. Traditional preprocessing techniques often require dropping features with many missing values or converting categorical variables into high-dimensional one-hot encodings, which can lead to information loss and reduced data utility. Moreover, such transformations may obscure the clinical meaning of the data, particularly in the context of surveys with unstructured or multi-option responses. Our method mitigates these limitations by converting only the available patient-specific information into natural language text, which is then embedded using an LLM pre-trained on medical datasets. This allows for more comprehensive data utilization without the need to explicitly manage missing values or predefined categorical schemas. Additionally, the LLM can infuse domain knowledge into the embedding vectors, enhancing their clinical relevance and semantic richness. Furthermore, by grouping patient information by clinical theme (e.g., symptoms, medications, comorbidities) and generating separate embeddings, the method improves the interpretability of downstream tasks. The proposed framework enables robust, knowledge-informed feature representation from medical data, facilitating more accurate and explainable machine learning in real-world healthcare applications.

2 - Investigating Grey-White Matter Contrast Changes In Healthy Older Adults

Jonas Swope, Oklahoma State University, Stillwater, OK, United States, Mohammad Fili, Guiping Hu, Auriel Willette

Gray-white matter contrast (GWC) reflects the clarity of the boundary between gray and white matter and serves as a key marker of brain health. In this study, we analyzed regional changes in GWC using structural MRI data from the UK Biobank. Participants were categorized into two groups based on substantial increases or decreases in GWC using the Desikan-Killiany Freesurfer brain atlas. These two groups were named: “Contrast Gainers” and “Contrast Decliners”. We computed Pearson correlation coefficients between brain regions and employed Jennrich’s test to evaluate statistical differences in correlation structures between the two groups. The results revealed significant differences between the correlation matrices of Contrast Gainers and Decliners. Heatmaps highlighted clusters of strongly correlated brain regions within each group, and further analysis uncovered both shared and distinct correlation patterns between them.

Keywords: Cognitive aging, Correlation patterns, Grey-white matter contrast, Structural Magnetic Resonance Imaging (sMRI).

3 - Multimodal Learning for Embryo Viability Prediction in Clinical IVF

Yicong Li, Harvard University, Cambridge, MA, United States, Junsik Kim, Zhiyi Shi, Davin Jeong, Johannes Knittel, Helen Yang, Yonghyun Song, Wanhua Li, Dalit Ben-Yosef, Daniel Needleman, Hanspeter Pfister

In clinical In-Vitro Fertilization (IVF), identifying the most viable embryo for transfer is important to increasing the likelihood of a successful pregnancy. Traditionally, this process involves embryologists manually assessing embryos’ static morphological features at specific intervals using light microscopy. This manual evaluation is not only timeintensive and costly, due to the need for expert analysis, but also inherently subjective, leading to variability in the selection process. To address these challenges, we develop a multimodal model that leverages both time-lapse video data and Electronic Health Records (EHRs) to predict embryo viability. One of the primary challenges of our research is to effectively combine time-lapse video and EHR data, owing to their inherent differences in modality. We comprehensively analyze our multimodal model with various modality inputs and integration approaches. Our approach will enable fast and automated embryo viability predictions in scale for clinical IVF.

4 - Optimizing the Johns Hopkins Fall Risk Assessment Tool (JHFRAT) for Improved Fall Risk Identification

Emmett Springer, Johns Hopkins University, Baltimore, MD, United States, Fardin Ganjkanloo, Erik Hoyer, Daniel Young, Kimia Ghobadi

This study aims to enhance the predictive performance of the Johns Hopkins Fall Risk Assessment Tool (JHFRAT) using electronic health record data and advanced analytics. A retrospective analysis was performed on 54,209 inpatient admissions from three Johns Hopkins Health System hospitals between March 2022 and October 2023. Patients were labeled high-risk or low risk for falls based on the occurrence of a fall during hospitalization and the number of targeted fall prevention interventions received. JHFRAT scoring was refined using a mixed-integer optimization problem that maximized a weighted sum of sensitivity and specificity while adhering to clinical knowledge constraints. The MIP model was compared against a constrained logistic regression and XGBoost models, which also incorporate additional EHR variables. All model performances were validated using stratified 5-fold cross-validation. The optimized JHFRAT model accurately identified 79.1% of high risk patients, a substantial

improvement over the original JHFRAT's sensitivity of 31.7%, while maintaining a specificity of 93.7%. The area under the precision-recall curve increased from 0.44 for the original JHFRAT to 0.58 for the optimized model. The analysis highlighted the critical importance of impaired mobility and cognitive impairments in predicting falls. By controlling for fall prevention interventions and maximizing sensitivity and specificity, the accuracy and predictive performance of JHFRAT risk stratification were substantially improved. This evidence-based approach provides a robust foundation for health systems to systematically enhance inpatient fall prevention protocols and patient safety using data-driven optimization techniques, contributing to improved risk assessment and resource allocation in healthcare settings.

5 - Targeted Priority Mechanisms in Organ Transplantation

Ruochen Wang, Virginia Tech, Blacksburg, VA, United States, Sait Tunc, Burhaneddin Sandikci, Bekir Tanriover, Matt Ellis

The persistent imbalance between organ supply and demand poses a significant challenge for the life-saving treatment of transplantation. This study introduces innovative targeted priority mechanisms inspired by the Eurotransplant Senior Program (ESP), designed to bridge the supply-demand gap while optimizing the matching between organs and recipients without mandating offer acceptance. These mechanisms diverge from traditional prioritization by offering voluntary participation and granting explicit priority to disadvantaged patients for a specified range of organs provided they restrict their offer acceptance to this set. Using a comprehensive queueing model, our study analyzes the strategic decisions of waitlisted candidates under targeted priority mechanisms in equilibrium, establishing the optimal program design. Our findings reveal the capacity of these mechanisms to separate organ allocation across different patient classes, improving overall system efficiency. Our evaluation of these mechanisms' impact on social welfare underscores their potential to improve overall welfare without adversely affecting any patient group. The practical applicability and impact of these mechanisms are demonstrated through a case study within the U.S. kidney allocation system using a clinically detailed simulation model, targeting elderly patients aged 65 and above. Simulation outcomes reveal a significant participation rate even when the program offers priority over kidneys that are conventionally perceived as marginal. Furthermore, our findings highlight the mechanisms' potential to increase yearly transplants by 1,000 and prevent up to 200 waitlist deaths annually. Improvements in the outcomes of even nonelderly patients highlight the mechanisms' potential to comprehensively address organ underutilization and improve organ-recipient matching in a real-world context.

SB41

Building B Level 3 B301

Bridging Research and Practice: The Role of Innovation, Regulation and Partnerships

Panel Session

Health Applications Society

Co-Chair: Kanix Wang, University of Cincinnati, Cincinnati, OH, United States

1 - Moderator Panelist

Kanix Wang, University of Cincinnati, Cincinnati, OH, United States

This year's panel will address challenges of integrating research into healthcare practice, focusing on interoperability, stakeholder alignment, and compliance. Discussions will emphasize the importance of multidisciplinary partnerships across academia, industry, and regulators, highlighting effective collaboration models to scale innovations. Additionally, the panel will explore how advancements in digital health, machine learning, and modeling influence clinical practices and shape future healthcare policies.

2 - Panelist

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

3 - Panelist

Turgay Ayer, ISyE Georgia Tech, Atlanta, GA, United States

4 - Panelist

Tolga Tezcan, Rice University, Houston, TX, United States

5 - Panelist

Ramesh Johari, Stanford University, Stanford, CA, United States

SB42

Building B Level 3 B302

Empirical Healthcare Studies

Invited Session

Health Applications Society

Chair: Bob Batt, Wisconsin School of Business, UW-Madison, Madison, WI, United States

1 - Machine Learning-Guided Cancer Screening: the Benefits of Proactive Care

Carri Chan, Columbia Business School, New York, NY, United States, Minje Park, Keith Boell, Elliot Mitchell, Abdul Tariq, David Vawdrey

With the advance of data analytics, many disease prediction models have been developed with the intent of detecting diseases earlier and improving patient outcomes through earlier treatment. The operationalization of interventions and care based on these predictive models is critical to attaining these goals. We study the real-world effects of a machine learning-guided colorectal cancer screening program deployed at a health system in Pennsylvania. Using a regression discontinuity design based on the predicted risk score for having cancer, we find that the program increases the likelihood of colonoscopy uptake in three and six months by 6.0 percentage points (214% increase relative to the control sample within the bandwidth) and 6.9 percentage points (117% increase), respectively. Importantly, we also find significant effects on mortality. We estimate that the program decreases 2-year mortality by 6.2 percentage points (43% decrease). Our finding suggests that a machine learning-guided proactive cancer screening program could significantly improve patient outcomes in addition to achieving higher disease detection rates. We argue that establishing unbiased estimates of the impact of machine learning-guided screenings is critical for capacity planning of screening resources, such as colonoscopies.

2 - A Tale of Two Eds: Demand-Inducing Effects of Opening New Emergency Departments

Hyun Seok Lee, Korea University Business School, Seoul, Korea, Republic of, Eric Park, Bob Batt, Timothy Rainer

Timely access to emergency care depends on both geographic proximity and wait times. While opening a new emergency department (ED) improves access, it may also induce demand by lowering the cost—time and effort—of seeking care. This study examines the demand-inducing effects of ED openings in two distinct contexts: a high-capacity urban area and an underserved rural region. Using comprehensive ED visit data from Hong Kong, we assess changes in ED visits within 5 km of newly opened facilities.

Overall, ED visits per district increased by 28.3% near the new urban ED and by 272.5% near the rural one. The impact was smaller in areas that already had nearby EDs before the new opening. The increase was most pronounced among younger populations. In the rural area, visits rose by 511.0% for individuals aged 0–19, declining progressively with age: 314.1% (20–39), 282.3% (40–59), 234.0% (60–79), and 103.4% (80+). A similar monotonic decreasing pattern by age was observed in the urban area, with a 47.6% increase for the youngest group and no significant change for the oldest. Low-acuity visits (triage levels 4–5) surged—by 44.8% in the urban and 406.3% in the rural area—while high-acuity visits (levels 1–2) increased more modestly.

These findings show that new ED openings substantially increase utilization—particularly among younger, lower-acuity patients—highlighting the need to account for induced demand in emergency care planning.

3 - From Delay to Disparity: Ethnic Gaps in Emergency Care Access and Outcomes Under a Universal Public Health System

Eric Park, Wake Forest University, Winston Salem, NC, United States, Timothy Rainer

This study examines racial and ethnic disparities in access to emergency medical care and the quality of care within Hong Kong's universal public healthcare system. Using 4.5 million patient-level emergency department (ED) visit records from 2013 to 2017, we analyze whether patients of different ethnic backgrounds experience unequal waiting times to be seen by a physician. We focus on five major non-Chinese ethnic groups—Pakistani, Filipino, Nepalese, Indian, and Indonesian—who represent the largest non-Chinese patient populations in the dataset.

Controlling for a wide range of patient, temporal, clinical, and hospital-level covariates, we find that these five ethnic groups consistently experience significantly longer waits compared to native Chinese patients, particularly among lower-priority triage levels (III–V). No disparity is observed among triage levels I and II, where clinical urgency mandates immediate or rapid attention. The observed differences suggest that racial and ethnic minority patients, often from marginalized labor sectors, may face systemic delays in receiving care even within a nominally equitable system.

These findings raise concerns about hidden barriers to equitable emergency care in Hong Kong's public hospitals and highlight the need for further investigation into potential sources of bias or structural inequities in ED operations. Ongoing work will examine differential rates of treatment abandonment by ethnicity to assess further implications for health outcomes.

4 - Early Birds Versus Last-Minute Arrivals: Empirical Evidence and Theoretical Analysis of Arrival Time Queueing Game

Xiangjie Zhao, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, Yichuan Ding, Dongdong Ge, Xiaoqing Xie

We examine a scenario where each customer strategically selects their arrival times at a queue, aiming to minimize their total costs, which include waiting, tardiness, and earliness costs. Building on the mixed strategy framework characterized by Haviv(2013), which provides a closed-form arrival time distribution, our focus extends to the transient behavior of this system in presence of irrational customers, in which case it is not yet known whether an equilibrium exists or whether the system would converge to the equilibrium. We introduce a dynamic choice model to describe customers' behavior in a repeated game setting. We empirically validate this model using data from a large endoscopy department in China, operating on a first-come first-served basis. The empirical result supports the efficacy of this model in capturing the adaptive strategies of patients concerning their arrival times. Leveraging this model, we prove that there exists a unique equilibrium and the system converges to this equilibrium under both a fluid model and an M/M/1 queueing model. By analyzing the equilibrium arrival time distribution, we identify the optimal time for patients to arrive, which occurs either at the start or near the end of the morning session, depending on the model parameters. Furthermore, we formulate a dynamic control problem to characterize the optimal information sharing policy, suggesting that service providers may decrease social cost by deliberately modulating the communication of waiting time information to patients. Further examination of social cost unveils that increasing earliness cost can, counter-intuitively, detract from social cost at equilibrium.

SB43

Building B Level 3 B303

Data-Driven Decision Making in Healthcare and Public Health

Invited Session

Health Applications Society

Chair: Hussein El Hajj, Santa Clara University, Santa Clara, CA, 95053, United States

1 - Identifying Regions Vulnerable to Obstetric Unit Closures Using Facility Location Modeling with Patient Behavior

Meghan Meredith, Georgia Institute of Technology, Atlanta, GA, United States, Lauren Steimle

The U.S. has the highest maternal mortality rate among high-income countries. This is just the tip of the iceberg with many more women experiencing severe complications due to their pregnancy. One commonly cited challenge in maternal healthcare delivery is that many pregnant people lack access to obstetric care, and this has been exacerbated by widespread closures of obstetric units. In this talk, we present a framework for identifying regions that are vulnerable to obstetric unit closures using facility location modeling. When considering the impacts of closures, we incorporate patient care-seeking behavior as many pregnant people "bypass" their closest obstetric facility to deliver at a farther facility. Altogether, we consider a problem in which an adversary seeks to maximize a population's average distance traveled to obstetric care by closing obstetric units. After the adversary closes obstetric units, patients probabilistically select the facility that maximizes their utility under a random utility maximization framework. We demonstrate that this model can be represented as a mixed-integer nonlinear program. Unlike existing facility location models that incorporate consumer choice, our model's objective function is not concave in general and thus, the state-of-the-art algorithms for solving facility location models with choice are not applicable. We introduce reformulations of our problem that lend themselves to solution methods which can take advantage of their decomposable structure. We compare our solution methods in numerical experiments and illustrate our framework using a case study of obstetric care in the state of Georgia.

2 - Navigating Incentive Challenges in Capitation Payment Models

Tolga Tezcan, Rice University, Houston, TX, United States

Healthcare systems are increasingly adopting capitation payment models to reduce costs. In the conventional capitation payment model, insurance companies receive predetermined, risk-adjusted payments per patient to provide care over fixed time periods, regardless of actual healthcare utilization. Although this model encourages insurers to enhance cost efficiency in theory, the actual outcomes have not always aligned with this expectation, as evidenced by the increased healthcare costs experienced by Medicare following the introduction of Medicare Advantage Plans.

Motivated by this observation, we demonstrate, using a simple model, that the conventional capitation model may not effectively incentivize cost reduction in competitive insurance markets when patients can enroll in any available plan. This failure stems from insurers exploiting imperfections in risk adjustment to cherry-pick profitable patients by manipulating plan designs, especially when patients have the option to enroll in alternative plans like the fee-for-service-based traditional Medicare Plan. Additionally, we show that, even with perfect information about patient types, which eliminates patient selection, conventional capitation may not yield socially optimal outcomes due to externalities stemming from patient choice.

Next, we introduce an innovative retrospective capitation payment model that leverages actual healthcare utilization of each patient. This approach eliminates the need for risk-adjustment procedures and the requirement to predict future medical expenses, all while still incentivizing providers to enhance cost efficiency. Using a general model, we demonstrate that this approach yields identical outcomes to the conventional capitation model with perfect patient type information in a competitive equilibrium.

3 - Towards Enabling the Last-Mile Logistics of Vaccine Supply Chains: Learnings from COVID-19 Vaccine Administration

Bhupinder Juneja, Carlson School of Management, University of Minnesota, Minneapolis, MN, United States, Ujjal Mukherjee, Kingshuk Sinha

Vaccine coverage can minimize the impact of the pandemic. However, there is a wide spatial disparity in vaccine coverage. While the significance of vaccine hesitancy and associated socio-economic factors are well-acknowledged, the factors related to vaccine delivery and administration from a supply-chain efficacy perspective are poorly understood. We analyze the effect of the enablers of the last mile of vaccine supply chains along 3-dimensions: logistical connectivity, delivery infrastructure, and informational connectivity. Our results suggest that all three dimensions of the last-mile delivery and administration of vaccines improve vaccine coverage. Further, we

demonstrate that improving enablers of the last mile of the vaccine supply chain is more beneficial for locations with relatively higher prevalence of socio-economic and demographic risk factors.

4 - A Patient Prospective on Treatments for Substance use Disorder

Zlatana Nenova, Wake Forest University, Winston-Salem, NC, United States, David Rea, Margret Bjarnadottir, Madeline Marks

We collected data from 578 U.S. residents who self-identified as having received individual therapy for alcohol use disorder. Across four surveys, we gathered background information as well as detailed data on their treatment experiences, personal goals, prior treatments, and perceived successes and failures. Our primary goal is to identify features of rehabilitation programs that help individuals with substance use disorder regain control over their lives. Through our analysis, we uncover major drivers of post-discharge success, including - but not limited to - group dynamics, therapy methodologies, and internal motivational factors. Our findings have the potential to help treatment centers better understand which practices, when properly reinforced, can significantly improve clients' chances of long-term success after discharge.

5 - Better Resource Allocations in the Criminal Justice System: Optimizing for the Long-Term Good

Chuwen Zhang, The University of Chicago Booth School of Business, Chicago, IL, United States, Pengyi Shi, Amy Ward

Incarceration diversion programs provide an alternative to incarceration, with a focus on rehabilitation over punishment. The goal is to prevent reoffense by providing treatment tailored to individual needs (for example, substance abuse problems). The hope is that such programs move society to a better place in the long term, with a smaller percentage of the population turning to crime. The question is: How do we decide which individuals are best suited for such programs?

To answer the question, we develop a framework that begins with survival models of individual reoffending behavior, incorporating both treatment effects and community-level interactions. We then derive a macro-level dynamical system that captures the mean behavior of individuals across different cohort groups. Theoretically, we establish uniform convergence of this dynamical system to a unique equilibrium under mild conditions on stationary policies. This allows us to study the effect of treatment policy decisions in the long term, and to highlight the inadequacy of myopic policies. We use a publicly available dataset together with an agent-based simulation to calibrate our dynamical system model, and validate our insights.

SB44

Building B Level 3 B304

Innovations in Healthcare Service and Research

Invited Session

MSOM: Healthcare

Chair: Miao Bai, University of Connecticut, Storrs, CT, United States

1 - Outpatient Appointment Scheduling with Waiting TIME Limits

Rachel Chen, University of California at Davis, Davis, CA, United States, Han Zhu, Min Zhang, Zhenzhen Zhang

The paper studies an outpatient appointment scheduling problem with waiting time limits, in the presence of uncertain service times, patient no-shows and unpunctuality. The presence of waiting time limit amplifies model complexity by significantly increasing problem scenarios. To address this challenge, we introduce the concept of virtual waiting time, representing the additional waiting time that a patient would have to wait to see the doctor after the waiting time limit is reached. Using this concept, we are able to unify the modeling of system dynamics of all different scenarios into one stochastic program. We develop a tailored integer L-shaped method to solve this model. We find that the presence of waiting time limits increases the job allowance between two adjacent patients, and the optimal schedule does not necessarily exhibit the well-known dome-shaped pattern.

2 - Sharing in Caring: Viewing Shared Medical Appointments Through An Operations Lens

Sundara Natarajan Panchanatham, HKU Business School, Hong Kong, Hong Kong, Enver Yucesan

Shared medical appointments (SMA), doctor-patient visits in which groups of patients are seen by one or more healthcare professionals in a concurrent session, are designed to achieve a more efficient use of medical resources and offer quicker –and less costly– access to healthcare services for patients. Moreover, for certain conditions, SMA can improve compliance and improve outcomes due to the learning effects from the peer group and the social benefits of meeting a group. Thus, the setup can potentially enhance patient experience, eliminate redundant work, and reduce costs for specific treatments. Our objective is to quantify the impact of SMA models on the cost and benefits for physicians and patients and characterize conditions under which SMA is a superior healthcare delivery model compared to the traditional care delivery models. On the methodology front, we mathematically model the operational trade-offs in SMAs and derive optimal capacity decisions for the healthcare providers using a game-theoretic framework.

3 - An Empirical Investigation of R&D Strategy and Clinical Trial Design

Kanix Wang, University of Cincinnati, Cincinnati, OH, United States, Zhaowei She, Liang Xu

This study examines the impact of eligibility criteria in clinical trials on drug approval, effectiveness, and inclusivity. Using AI-driven models, we analyze how criteria evolve across trial phases, their influence on FDA approval and post-market outcomes, and the role of firm size in balancing trial success with broader patient population.

4 - Dynamic Multi-Class Multi-Appointment Advance Scheduling: Approximate Policies and Lower Bounds

Vahid Sarhangian, University of Toronto, Toronto, ON, Canada, Yanbo Ding

We study a dynamic advance scheduling problem with multiple classes of customers and with each customer requiring several appointments. Our primary motivation is scheduling appointments for Radiation Therapy (RT) treatments for patients with different urgency levels which usually consist of multiple appointment sessions. We consider a Markov Decision Process (MDP) formulation of the problem and propose scalable approximate dynamic programming policies based on hindsight optimal policies. We evaluate the performance of the proposed policies using a related information-relaxation lower bound, and in comparison with other benchmarks. Finally, we conduct a case study using real data to illustrate the effectiveness and practicality of the proposed policies.

SB46

Building B Level 3 B306

Sequential Search and Information Design

Invited Session

Decision Analysis Society

Chair: Zhenyu Hu, National University of Singapore, Singapore, Singapore

Co-Chair: Shouqiang Wang, University of Texas Dallas, Richardson, 75080, United States

1 - Why the Best Machine May Not be the Best: Incentivizing Human-Machine Collaboration

Xiaotong Guan, The University of Texas at Dallas, Richason, TX, United States, Anyan Qi, Shouqiang Wang

Humans and machines need to work collaboratively to make operational decisions such as demand forecasts. In this paper, we study how the machine should be designed to incentivize human-machine collaboration and obtain a rich set of results.

2 - Persuasive Calibration

Wei Tang, Chinese University of Hong Kong, Hong Kong, Hong Kong, Yiding Feng

We introduce and study the persuasive calibration problem, where a principal aims to provide trustworthy predictions about underlying events to a downstream agent to make desired decisions. We adopt the standard calibration framework that regulates predictions to be unbiased conditional on their own value, and thus, they can reliably be interpreted at the face value by the agent. Allowing a small calibration error budget, we aim to answer the following question: what is and how to compute the optimal predictor under this calibration error budget, especially when there exists incentive misalignment between the principal and the agent? We focus on standard ℓ_1 -norm Expected Calibration Error (ECE) metric. We provide both algorithmic solutions and structural characterizations for optimal calibrated predictors.

3 - Best is Not Always First: Optimizing Offer Deadlines and Offer Sequence with Delayed Feedback

Wenjie Tang, University of Vienna, Vienna, Austria, Zhenyu Hu

We study a setting where a proposer (e.g., an employer) with a single offer to make sequentially approaches a set of responders (e.g., job candidates) each of whom is also searching for alternative offers. We are interested in the proposer's decision of which responder to offer to first, and the optimal deadline to set for the responder's reply. We provide a full characterization of the optimal deadlines. When the proposer's choice set is composed of only two responders, we also fully characterize the proposer's sequence decision: She should offer to the higher-utility responder first if and only if she approaches the responders later than a time threshold. Moreover, we show that, among exploding responders who are able to provide immediate feedback on the proposer's offer, the proposer should always prioritize the highest-utility one—as long as responders are homogeneous in the probabilities of their availability at any given time. In addition, when all but one of the responders are exploding responders, we show that as the proposer approaches this choice set later in the time horizon, the sole non-exploding responder should be prioritized over fewer exploding responders. Finally, our analysis of a static greedy policy that always prioritizes the highest-utility responder indicates that a greedy policy can be rather efficient when the market is not very competitive; otherwise, our numerical examples show that the proposer can incur substantial loss.

4 - Information Design of the Monty Hall Problem

Zhenyu Hu, National University of Singapore, Singapore, Singapore, Shouqiang Wang, Yangge Xiao

We consider the following variant of the Monty Hall Problem. There are n doors, behind each of which hides an a priori random reward, independently distributed according to a common prior. A clue box is attached to each door, which, if inspected, yields a clue about the rewards behind possibly all doors. The game host designs what clues to be put in each box and announces at the beginning of the game what each clue reveals about the rewards. Given the design, the player decides sequentially which clue box to inspect by paying an inspection fee and when to stop inspecting the boxes. Upon stopping, the player can then utilize the clues collected so far to open a door whose attached box has already been inspected and take the reward behind it. The player wishes to maximize the reward obtained from the opened door, while the host maximizes the fees collected from inspections. Under specific designs, the player's problem can be reduced to several classic sequential information acquisition problems, such as the Pandora's box problem or the Wald's sequential sampling problem. We study the optimal designs of the clue boxes when the host faces different informational constraints. These informational constraints arise from the potential applications in different settings such as product-page design on online platforms, information design in delegated search, and reward design in video games.

SB47

Building B Level 3 B307

Platform Design

Contributed Session

Contributed

Chair: Tania Saha, S.P. Jain Institute of Management & Research (SPJIMR), Mumbai, India

1 - Shoulders of Little Giants: A Two-Stage Design to Assist Solution Providers in Innovation Contests

Sriram Sankaranarayanan, Indian School of Business, Mohali, India, Swanand Deodhar

Tournaments mediated by digital platforms have become a key mechanism for sourcing innovation, allowing seekers to solicit solutions from a distributed pool of providers. However, solution quality is uncertain, and platform mechanisms to assist providers often require greater seeker involvement, raising transaction costs. To address this, we propose a two-stage contest design. In the first stage, artifact-development (AD), providers create artifacts; in the second, solution-development (SD), they use these artifacts to develop final solutions. We model three types of artifacts: (i) expertise-enhancing, improving top providers; (ii) expertise-equalizing, aiding mid-and low-tier providers; and (iii) cost-reducing, lowering effort costs for all. We identify conditions for when this two-stage design offers better seeker return than a traditional one-stage design, and derive optimal split of the incentive budget between these two stages, resulting in meaningful takeaways for the seeker. For example, we show that when expertise-enhancing or expertise-equalizing artifacts are encouraged, it is optimal for them to have an AD stage and offer increasing larger share of reward to the AD stage, when the seeker expects a large participation in the SD stage.

2 - A Community Recharging Facility Sharing Program: Optimized Slot Assignment Considering Uncertain User Acceptance

Hongke Xie, University of Electronic Science and Technology of China, Chengdu, China, People's Republic of, Pengyu Yan

The rapid adoption of electric vehicles (EVs) has intensified the demand for accessible charging infrastructure, particularly in urban areas where private charging facility installation is often infeasible. While shared charging programs offer a potential solution, existing approaches suffer from low utilization due to mismatches between supply and demand, with only 1.57% of private charging facilities currently shared in China. This study proposes a novel community-based EV charging facility sharing program, where users with charging-equipped parking slots (C-users) temporarily exchange with users lacking such facilities (N-users) overnight. A key challenge is optimizing slot assignments without prior knowledge of users' walking distance-dependent acceptance probabilities, which are private and stochastic. To address this, we formulate the problem as a Sleeping Combinatorial Multi-Armed Bandit (CS-MAB) and develop a Sleeping Combinatorial Upper Confidence Bound (CS-UCB) algorithm that learns acceptance probabilities while maximizing fulfilled charging demands. Furthermore, we integrate a Structural Causal Model (SCM) framework to infer counterfactual feedback and accelerate learning during initial deployment stages. Numerical experiments demonstrate that our approach achieves near-optimal performance, with regret below 11.86% across varying demand-supply ratios, significantly outperforming benchmarks. Sensitivity analyses confirm robustness to stochastic user behavior, exploration parameter, and pricing schemes. The extension verifies one-to-one exchange mechanism proves more effective than many-to-many assignment, ensuring practical scalability and user satisfaction.

3 - Dynamic Random Bipartite Matching Under Spatiotemporal Heterogeneity for Mobility Services

Shiyu Shen, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Yanfeng Ouyang

This study explores a variation of the online bipartite matching problem that allows randomness in the spatiotemporal distribution of the vertices, referred to as the Spatio-Temporal Random Bipartite Matching Problem (ST-RBMP). This type of problem directly builds upon spatial location information of the vertices, and can be applied to services such as shared mobility (e.g., matching riders with drivers). Many strategies and algorithms have been developed to solve some form of ST-RBMP in different contexts. However, the inherent nature of dynamic and stochastic decision-making imposes notable challenges, such as the so-called wild goose chase (WGC) phenomenon. To address these challenges, this paper proposes a modeling framework to dynamically determine the optimal pooling intervals and matching radii for ST-RBMP to minimize the system-wide costs associated with matching distance and waiting for mobility services systems. At each matching epoch, the system state is represented by the spatial distribution of demand and supply points, referred to as a vertex distribution profile. Closed-form formulas for the expected matching distance and probability of static RBMPs are adapted into a time- and space- dependent decision framework, and now with consideration of max radii, to minimize the total expected matching distances. Following a continuum approximation scheme, the optimal functions of pooling intervals and matching radii over time and space are then derived from the Euler-Lagrange equations. This framework effectively addresses the spatiotemporal heterogeneity in demand and supply distributions in many real-world systems. Moreover, it is transferable and robust across diverse problem settings and varying decision-making scenarios.

4 - Exploring the Impacts of Mobilitycoins on Last-Mile Delivery Services

Anke Ye, Technical University of Munich, Munich, Germany, Maximilian Schiffer, Klaus Bogenberger

Logistics companies aim to improve user experience by offering a variety of personalized last-mile parcel delivery services. However, the additional mileage associated with these services can hinder efforts to achieve decarbonization goals. In order to promote sustainable ordering habits amongst users, we propose an innovative internalization of the services' environmental externalities by means of a Tradable Credit Scheme (TCS) known as MobilityCoins (MoCo). This work represents one of the first attempts to apply TCS in a logistics context, specifically focusing on the MoCo scheme design and its impacts on the logistics company and users. To derive tangible insights, we model the system using a three-layer programming approach. At the upper level, the MoCo agency optimizes the scheme design by determining the total MoCos in the initial allocation based on the predetermined per-kilometer MoCo charge. The middle level captures the pricing problem faced by logistics companies, which offer two service options, i.e., Express and Standard. These options differ with respect to delivery times and consolidation potential. A shift in demand from the faster service to the more energy-efficient alternative ultimately affects the amount of MoCos consumed for delivery. At the lower level, users with diverse price sensitivities choose their delivery service following the user equilibrium principles. We first investigate the system's features in a monopoly market and then extend the analysis to a competitive market with multiple logistics companies. The findings provide valuable managerial insights into the extended application of TCS.

5 - On Robust Inventory Optimization for Bike Sharing

Haixiang Wang, King's College London, London, United Kingdom, Dimitrios Letsios

We aim at enhancing service level in bike sharing systems under demand uncertainty. We develop a robust optimization approach for station-level inventory optimization. Although deterministic optimization models effectively evaluate past planning decisions with known demand, their performance is deteriorated when considering future bike allocations due to demand prediction errors. We consider demand realizations within Gamma uncertainty sets. We formulate a deterministic optimization model and express its robust counterpart as a bi-level optimization problem. Unlike previous inventory optimization approaches that rely on dualization techniques for computing robust solutions to the inventory rebalancing problem, we propose a novel approach based on Benders decomposition, including primal heuristic and adversarial relaxation components. The robustness of our approach is validated computationally.

6 - Slide or Feed: Content Recommendation with Creators' Incentives

Shu Zhang, the Chinese University of Hong Kong, Hong Kong, Hong Kong, Hongfan Chen, Renyu Zhang, Zhou Zhou

In online video platforms, optimizing recommendation systems is crucial for maximizing profit and incentivizing content creators. Our work delves into the comparative efficacy of two layout modes of recommendation systems in online content platforms: the single-column (slide) and double-column (feed) formats. We build up a gaming model to describe the dynamics between recommendation modes and their impact on platform economics, where the two modes are different in both recommendation accuracy as well as their monetization ability.

We illustrate how the performance of recommendation system is affected by both the modes and the users' interactions on the platform. The equilibrium results show that the strength of the recommendation algorithm plays an important role in this comparison. Platforms with robust algorithms benefit from the slide mode's efficiency, while those with weaker algorithms should leverage the feed mode to enhance usability and engagement. Several operational factors of the platforms also interplay in this comparison.

Insights from this study advance our understanding of how recommendation system modes affect user content consumption patterns, creator engagement, and the financial sustainability of online content platforms. These findings also offer actionable guidance for platform managers and market regulators seeking to optimize operations and policy-making.

7 - A Manufacturer's Pricing Format in A Dual-Distribution Channel of Retail and Rental Platforms

Tania Saha, S.P. Jain Institute of Management & Research (SPJIMR), Mumbai, India, Prasenjit Mandal

Many manufacturers use a dual-distribution channel strategy, offering products for purchase through retail platforms and sharing products through rental platforms. For example, Tommy Hilfiger partnered with rental platform MY Wardrobe HQ and also with the retail platform Kohl. While retail platforms enable consumers to get exclusive ownership of products, rental platforms facilitate temporary access to products. In the presence of both retail and rental platforms, consumers have the option to buy or rent. Furthermore, manufacturers are not only involved in wholesaling their products (the traditional wholesale pricing format) but also using retail/rental platforms as marketplaces where they sell/share the products directly with the consumers (the agency pricing format). The manufacturer can opt for the same pricing format with both platforms, i.e., either a wholesale pricing format or an agency pricing format. The manufacturer can also opt for a hybrid pricing format approach, where the manufacturer opts for different pricing formats with the two platforms (for example, wholesale with the retail platform and agency with the rental platform or vice versa). In the wholesale pricing format, the retail (rental) platform purchases products from the manufacturer for a wholesale price and then decides the retail (rental) prices for consumers. In the agency pricing format, the firm determines the retail (rental) prices and shares revenue with the retail (rental) platform. In this paper, we study the manufacturer's pricing format: agency or wholesale with the retail and rental platforms?

SB48

Building B Level 3 B308

Data-driven Decision and Risk Analysis

Invited Session

Decision Analysis Society

Chair: Chen Mavis Wang, N/A

Co-Chair: Yifan Hong, Tsinghua University, Beijing, N/A

1 - Lamda: Large Language Models as Decision Analysts

Yifan Hong, Tsinghua University, Beijing, China, People's Republic of, Chen Wang

Influence diagrams help tackle the challenges of decisions under risk by structuring information, decisions and values, and explicitly representing uncertainties and probabilistic dependencies. However, constructing an influence diagram requires expertise in decision analysis and is further complicated by the need to process large amounts of contextual information. This work automates the construction of influence diagrams from natural language input with the help of large language models (LLMs). We design a workflow where LLMs are prompted to output elements of an influence diagram and resolve issues through verification and regeneration. We then construct a new dataset of typical decision problems under risk, and evaluations on this dataset demonstrate that our framework effectively identifies key factors and relationships in natural language and, therefore, makes better decisions than standalone LLMs and LLMs enhanced with standard techniques such as chain-of-thought (CoT). Furthermore, we extend the framework to aggregate opinions from multiple information sources, and apply the aggregation method to group discussions by disease control experts on a hypothetical pandemic scenario.

2 - Honesty, Deception, OR Secrecy? Optimally Disclosing Defensive Posture

Kyle Hunt, University at Buffalo, Buffalo, NY, United States, Sule Guney, Richard John

Due to the strategic and adaptive nature of adversaries, the deployment of new technologies is common practice in the arenas of security and defense. History has shown that the deployment of these technologies has often been disclosed to the public, allowing malicious actors to potentially understand which venues are defended, and how. Within this context, there is limited research examining how defenders should optimally disclose their deployment strategies -- a gap we fill in this work. To understand adversary decision-making in this setting, we first run behavioral experiments to understand how adversaries develop their attack plans in response to information

disclosed by a defender. Specifically, we address whether attackers (i) believe the information they receive from the defender and (ii) choose to attack after learning that new security measures may be deployed at their target(s) of interest. Then, utilizing the data from the experiments as a proxy for real adversary behavior, we optimize the disclosure strategy for the defender, where she can either be deceptive, secretive, or truthful. We compare these disclosure policies with those obtained when optimizing against different traditional models of adversary behavior that have long been adopted in the literature.

3 - The Role of Mental Simulation in Risk Decision-Making: Evidence from Air Traffic Management

Tong Deng, Tsinghua University, Beijing, China, People's Republic of

Effective risk decision-making is essential in high-stakes domains such as air traffic management (ATM), where safety outcomes depend on operators' ability to anticipate and mitigate evolving hazards. This study examines how mental simulation and environmental controllability jointly shape the causal basis of risk decisions among air traffic controllers. We develop an airspace simulation model to approximate controllers' mental simulations of future traffic states. Weather conditions serve as an exogenous proxy for environmental controllability, enabling a stratified analysis of decision-making under varying levels of environmental uncertainty. Using real-world operational data (including air traffic flows, weather, and controller operations), we uncover the underlying causal structure among air traffic risk features, controllers' mental simulations, and their risk decisions, and estimate the sizes of these causal effects. The results show that in good weather conditions with high controllability, mental simulation significantly contributes to decision-making, enabling proactive mitigation measures. In contrast, in bad weather conditions with low controllability, decisions are heavily influenced by static airspace structures and existing congestion, indicating an inclination toward reactive strategies. Our study presents a novel approach to leveraging real-world operational data and causal machine learning methods to elucidate the impact of simulation-based reasoning on safety-critical decision-making.

4 - Epidemic Mitigation Under Stochastic Contagion Models

Janne Kettunen, The George Washington University, Vienna, VA, United States, Miguel Lejeune

We propose a risk-averse prescriptive contagion model governed by stochastic time-varying transmission dynamics to guide epidemic mitigation. The model is a multi-period mixed-integer nonlinear chance-constrained stochastic programming problem that minimizes the burden of the mitigation actions whilst guaranteeing that the need for ICUs remains below their availability with high probability. The evolution of the pandemic is represented via a compartmental Susceptible-Exposed-Infectious-Recovered-Died model with a stochastic and mitigation-dependent transmission rate of the disease. We design a solution method that features the deterministic and equivalent reformulation of the nonlinear chance constraint using a Boolean method, quadratization of the set of nonconvex constraints using linearization methods for polynomial terms and exponential function with integer arguments, and construction of the envelope of bilinear terms defined on a specific domain with non-trivial upper bound on the product of the two variables. The proposed framework generalizes well to alternative problem formulations. Indeed, the reformulation is applicable to nonconvex polynomial chance-constrained problems outside the epidemic mitigation field. We apply the models to real COVID-19 data in four Nordic countries. The experiments confirm the computational efficiency and scalability of our approach and show that the solutions are stable, reliable and perform well out-of-sample, and robust even in the presence of uncertainties in other key model parameters. The results highlight the proper timing and stringency of mitigation actions. The optimal policy can include some alternations between periods of strict and lenient mitigation. Counterintuitively, mitigation can be optimal to forgo at early periods.

SB49

Building B Level 3 B309

Supply Chain Management for Social Good

Invited Session

MSOM: Supply Chain

Chair: Xiajun Pan, University of Florida, Gainesville, FL, United States

Co-Chair: Shiwei Chai, University of Florida, GAINESVILLE, FL, United States

1 - Operational-Marketing Trade-Offs in Managing Social and Environmental Responsibility

Shiwei Chai, University of Florida, Gainesville, FL, United States, Asoo Vakharia, Anand Paul

This paper is motivated by observing the substantive interest in environmental stewardship by both regulators and individual firms. For regulators, key concerns relate to better management of public environmental risks by designing an incentive mechanism (comprising a fine and a fixed subsidy) for a firm to optimize social environmental welfare. The firm, on the other hand, maximizes its profit by choosing product prices, marketing (advertising) effort, and operational (audit/inspection or process improvement) effort expended to reduce environmental violations.

Through a rigorous analysis of a stochastic dynamic programming problem for a risk-neutral firm and a characterization of the corresponding fine and subsidy mechanism for the regulator, insights offered are as follows. While the fixed subsidy has no impact on the firm's decisions, increases in the fine levied by the regulator incentivizes the firm to expend: (a) higher marketing effort through an increase in advertising expenditures; and (b) higher operational efforts directed at reducing the possibility of environmental violations. From the regulator's perspective, the average costs associated with an environmental violation serve as a lower bound on the fine that should be levied. Interestingly, we also find that when either the social and environmental cost or the loss of brand image due to environmental violations is above a certain threshold, the regulator should not only impose a fine but simultaneously provide a fixed subsidy to the firm.

2 - Boxless Shipping: Balancing Sustainability and Security in Package Delivery

Stanley Lim, Michigan State University, East Lansing, MI, United States, Neda Mirzaeian, Shouqiang Wang

We explore optimal packaging strategies for online retailers, focusing on boxless shipping to address environmental waste and theft concerns. Using an information design framework, we demonstrate profit gains of up to 221% and environmental benefits. However, we highlight that outcomes depend on the sensitivity of customers to theft and privacy.

3 - Product Rollover Strategies: Dynamic Decision Model Motivated by the Hybrid Seed Agribusiness

Jian Li, Northeastern Illinois University, Chicago, Illinois, IL, United States, Panos Kouvelis

We study the optimal launch timing and rollover pattern selection for new product's launch. We consider two product rollover patterns: Solo-Rollover and Dual-Rollover. We examine two distinct timing strategies: the Static approach, which ex-ante plans the launch time and/or rollover pattern of the new product, and the Dynamic approach, which dynamically determines these decisions accounting for realized demand information over time. Our focus is on formulating an optimization model that integrates the above-mentioned decisions in managing product rollovers over a given lifecycle. We also study, as special cases, five other timing and rollover strategies that differ along the static/dynamic nature of the launch's timing and contingent execution of the rollover pattern selection.

4 - A Model of Tokenized Marketplaces

Jiong Sun, Purdue University, West Lafayette, IN, United States, Shiwei Chai

We develop an analytical model to rationalize the economic value of digital tokens as a means of reducing reliance on intermediaries and improving transactional efficiency. Specifically, we propose a token-based framework for decentralizing digital marketplaces. In this setting, utility tokens function as a coordination mechanism that mitigates underproduction and protects producers from adverse price pressures typically exerted by conventional intermediaries. This is achieved through decentralized controls, where transaction validation is delegated to independent contributors and commission fees are determined by producers themselves. Our findings are robust across various contexts involving multiple producers, whether they offer unrelated, complementary, or substitute products.

SB50

Building B Level 3 B310

Overcoming the Challenges of Policy Optimization

Invited Session

MSOM: Service Operations

Chair: Hamsa Bastani, Wharton School, Philadelphia, PA, United States

Co-Chair: Bryce McLaughlin, University of Pennsylvania, Philadelphia, United States

1 - Inference-Aware Policy Optimization

Bryce McLaughlin, University of Pennsylvania, Philadelphia, PA, United States, Hamsa Bastani, Osbert Bastani

The traditional goal in machine learning is to build predictive models, which forecast outcomes that can be used to guide decision-making. When we attempt to optimize a decision-making policy using these models, it is hard to overcome the challenge that the model's accuracy varies significantly over the population. A naive approach that takes the decision associated with the best counterfactual outcome prediction will suffer as it will often act on noise, generating large variance in its observed performance.

This variance makes validating the policy using a hold-out test set extremely difficult, often preventing implementation of the policy. Essentially whatever estimate of policy improvement we produce will be yielded insignificant as the overconfident decisions made by some parts of the model drown out gains elsewhere. In settings where the outcomes following decisions are high-stakes, significant improvements are often needed to change the implemented policy, making the naive approach much less useful.

Our approach involves updating the objective of policy optimization from expected reward to a tradeoff between increasing expected reward and decreasing the variance induced by the policy. Our solution, Inference-Aware Policy Optimization (IAPO), traces out a Pareto frontier of policies, allowing the user to tradeoff between expected performance and stability. Using a large dataset of behavioral health nudges, we show that IAPO can identify and implement significant policy improvements while other common approaches deteriorate the policy almost as often as they improve it.

2 - Mind the Gap: Encouragement Policy Learning Balancing Take-up and Efficacy

Angela Zhou, University of Southern California, Los Angeles, CA, United States

In consequential domains, it is often impossible to compel individuals to take treatment, so that optimal policy rules are merely suggestions in the presence of human non-adherence to treatment recommendations. Under heterogeneity, covariates may predict take-up of treatment and final outcome, but differently. While optimal treatment rules optimize causal outcomes across the population, access parity constraints or other fairness considerations on who receives treatment can be important. For example, in social services, a persistent puzzle is the gap in take-up of beneficial services among those who may benefit from them the most. We study causal identification and robust estimation of optimal treatment rules, including under potential violations of positivity. We consider constraints such as demographic parity in treatment take-up, among others, via constrained optimization. Our framework can be extended to handle algorithmic recommendations under an often-reasonable covariate-conditional exclusion restriction, using our robustness checks for lack of positivity in the recommendation. We develop a two-stage algorithm for solving over parametrized policy classes under general constraints to obtain variance-sensitive regret bounds. We illustrate the methods in three case studies based on data from reminders of SNAP benefits recertification, randomized encouragement to enroll in insurance, and from pretrial supervised release with electronic monitoring. While the specific remedy to disparities in algorithmic allocation is context-specific, it requires studying both take-up of decisions and downstream outcomes of them.

3 - Designing AI-Human Supervision to Improve Worker Performance: a Field Experiment in Service Operations

Bing Bai, McGill University, Montreal, QC, Canada, Xueming Luo, Zheng Fang, He Peng

We study how firms can effectively integrate AI into supervisory roles to enhance worker performance. Using data from a randomized field experiment at a fintech firm, we compare the effects of different human-AI supervisory configurations on customer service agents. Workers received performance feedback from either human managers (human-solo), an AI bot (AI-solo), both in parallel (dual), or an AI system whose feedback was relayed by a human manager (shadow-AI). We find that AI-solo supervision improves worker performance relative to human-solo, but that combining both in a dual configuration backfires—leading to confusion, reduced feedback-seeking, and lower task performance. In contrast, the shadow-AI configuration improves performance by encouraging proactive learning and reducing customer frustration. Importantly, while AI-solo supervision exacerbates gender disparities, shadow-AI reduces the gap by improving performance for both genders. Our findings highlight that AI should not replace human managers, but rather work behind the scenes to support them, enabling more productive, equitable, and human-centered service operations.

4 - Large-Scale Personalization: Lessons from Behavioral Nudging in India

Divya Singhvi, New York University, New York, NY, United States

In this talk, I will share insights from recent work on designing and deploying a personalized behavioral nudging system to optimize user engagement at scale on an EdTech platform in India. I will present key findings from field experiments and highlight practical lessons for implementing personalization strategies in data-scarce environments. The discussion will focus on both methodological innovations and operational challenges, offering actionable guidance for researchers and practitioners working in similar contexts.

SB51

Building B Level 3 B311

Advances in Empirical Platform Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Sidika Candogan, National University of Singapore, Singapore, Singapore

Co-Chair: Guang Cheng, National University of Singapore, Singapore, Singapore

1 - Platform Entry in Winner-Take-All Markets: An Empirical Investigation of Feature-Based Differentiation of a Microblogging Platform

Jonathan Gomez Martinez, Marshall School of Business - University of Southern California, Atlanta, GA, United States, Ramnath Chellappa, Emma Zhang

Theoretical work on platform competition emphasizes the importance of differentiated features and multihoming in determining market outcomes. However, empirical evidence of platforms competition in stable markets remains sparse. We analyze such a market through the introduction and growth of the Parler social media platform. Parler, offered similar features to other microblogging platforms with the addition of freedom from content moderation as its differentiated feature. Our work demonstrates that Parler's differentiated feature, both, expanded the market and increased competition with the dominant microblogging platform. We also identify periods when the salience of Parler and its differentiated feature were heightened. We find that competition was heightened with additional salience of Parler's differentiated feature while increased visibility of Parler, itself, resulted in market expansion. Our work concludes with a discussion of the events leading to Parler shutting down and the competitive issues therein.

2 - The Role of Contextual Information in Customer Reviews: a Field Experiment on a Mental Health Platform

Guang Cheng, National University of Singapore, Singapore, Singapore, Sidika Candogan, Joel Goh, Bilal Gokpinar

Online mental health platforms can increase the accessibility of mental health care by allowing customers to select a therapist from a large pool of vetted professionals. However, customers may find it difficult to choose a therapist best suited to their needs because therapists typically list multiple specializations that tend to overlap significantly. This makes it challenging for online mental health platforms to attract new customers and retain existing ones. Motivated by this challenge, we conducted a randomized field experiment in partnership with Singapore's largest online mental health platform. In this experiment, we empirically investigate whether a simple intervention on the customer reviews—revealing contextual information about the review writers—can improve the platform's ability to attract and retain customers. We

find that, relative to the control, treated customers were 15% more likely to book at least one therapy session, and they booked 36% more sessions in total during the experiment. Furthermore, treated customers who have the first therapy session booked 55% more follow-up sessions with their first chosen therapist. Overall, our results suggest that revealing contextual information about the review writers improves the informativeness of customer reviews on mental health platforms, and it is a simple and effective way to increase successful matches and foster long-term relationships between customers and the platform.

3 - 30 Million Canvas Grading Records Reveal Widespread Sequential Bias and System-Induced Surname Initial Disparity

Jiaxin Pei, Stanford University, Palo Alto, CA, United States, Zhihan Wang, Jun Li

he widespread adoption of learning management systems in educational institutions has yielded numerous benefits for teaching staff but also introduced the risk of unequal treatment towards students. We present an analysis of over 30 million Canvas grading records from a large public university, revealing a significant bias in sequential grading tasks. We find that assignments graded later in the sequence tend to (1) receive lower grades, (2) receive comments that are notably more negative and less polite, and (3) exhibit lower grading quality measured by post-grade complaints from students.

Furthermore, we show that the system design of Canvas, which pre-orders submissions by student surnames, transforms the sequential bias into a significant disadvantage for students with alphabetically lower-ranked surname initials. These students consistently receive lower grades, more negative and impolite comments, and raise more post-grade complaints as a result of their disadvantaged position in the grading sequence. This surname initial disparity is observed across a wide range of subjects, and is more prominent in social science and humanities as compared to engineering, science and medicine. The assignment-level surname disparity aggregates to a course-level surname disparity of students' GPA and can potentially lead to inequitable job opportunities. For platforms and education institutions, the system-induced surname grading disparity can be mitigated by randomizing student submissions in grading tasks. Education institutions should keep the workload of graders at a reasonable level to reduce fatigue and/or have multiple graders as a cross validation to enhance grading quality.

4 - Have You Walked in My Shoes? Mitigating the Cold Start Problem in Peer-to-Peer Platforms with Bilateral Reviews

Junghyun Maeng, National University of Singapore, Singapore, Singapore, Varun Karamshetty, Hyeokkoo Eric Kwon

Early-stage sellers on peer-to-peer (P2P) platforms often struggle to attract buyers due to the cold-start problem—a lack of performance feedback reviews that limits their ability to establish trust with prospective buyers. This study investigates whether and how bilateral review systems can mitigate this problem by enabling the transfer of reputation between roles. Using a unique panel dataset of Airbnb property listings in New York City from 2015 to 2019, we exploit the variation in the timing of when new hosts receive guest-side reviews (reviews they earned in their role as guests) to estimate the impact on their performance as a host (seller). We find that receiving guest-side reviews increases the monthly occupancy rates of early-stage hosts and the number of booked days by 11.2% and 13.6%, respectively. Interestingly, these effects are stronger for early-stage hosts who face higher baseline uncertainty in performance, such as belonging to racial minority groups or operating in high-crime neighborhoods. We also find that the positive effect on host performance is amplified when guest-side reviews contain host-relevant content, such as feedback on their cleanliness and communication (as a guest). Our findings contribute to the literature on bilateral review systems and cold-start problem by showing that trust in role-fluid P2P environments is not siloed – reputation earned as a buyer can meaningfully shape seller outcomes. Based on our findings, we offer actionable insights for early-stage sellers and P2P platforms seeking to design bilateral review systems.

SB52

Building B Level 3 B312

Emerging Topics in Sustainable OM

Invited Session

MSOM: Sustainable Operations

Chair: Tim Kraft, North Carolina State University, Raleigh, NC, United States

1 - Getting to the Green: Should a Profit-Maximizing Firm Buy Carbon Offsets OR Invite Consumers to Buy Them?

Gokce Esenduran, Purdue University, West Lafayette, IN, United States, H. Sebastian Heese, Gil Souza

Many companies, including American Airlines, Patagonia, Google, and Apple have pledged to achieve carbon neutrality, pursuing initiatives such as renewable energy use, zero-emissions vehicles, and low-carbon materials. However, residual emissions often remain, requiring the purchase of carbon offsets, which raise costs. While environmentally conscious consumers may accept higher prices for low-emissions products, many climate-disengaged consumers will not. In this paper, we identify when it is optimal for a profit-maximizing firm to offer its consumers carbon-reduction offsets for purchase with the product, in addition to potentially purchasing offsets at the firm level. Through a stylized analytical model, we show that it is never optimal for firms to both buy offsets at the firm level and offer them to consumers for purchase simultaneously. Rather, when the offset cost is low enough, the firm buys enough offsets to compensate for all of its emissions at the firm level, resulting in a higher overall product price; when the offset cost is in the middle range, the firm offers offsets to consumers for purchase at an offset price lower than offset cost, ensuring that green consumers buy offsets; and when the offset price is high, offsets cannot be optimally used in any way. The thresholds depend on the green-segment size and disutility from emissions, and on the product's own carbon footprint. Providing offsets for purchase to consumers allows consumers to self-select into their preferred product/price bundle, thus providing the firm with a market-segmentation and price-discrimination mechanism that increases profits and reduces the firm's carbon footprint.

2 - Consumer Food Waste in Retail Operations

Seulchan Lee, Michigan Technological University, Hancock, MI, United States, Michael Ketzenberg, Han Oh, Nesim Erkip

Our research focus is on consumer food waste for sustainable retail operations. We specifically address the research question of how consumer food waste is affected by retail decision-making and the reciprocal, how food waste by consumers affects retail decision-making. We introduce a generic model of consumer utility with regard to food consumption and waste. From this granular level of heterogeneous consumer utility, we derive a stochastic, aggregate market demand function that drives retail inventory management. We do so in the context of a single period newsvendor problem. Several key observations arise from our analysis. For example, providing a higher quality product for sale may increase total waste. We also observe that increasing consumer disutility to waste may actually increase total food waste. In addition, we find that an increase in the unit salvage value of unsold inventory at the end of the period will also generally increase total waste. We contribute to the understanding of the behavioral mechanisms that explain how both food consumption and waste interact with and impact retailer decision-making. These findings have important implications with respect to potential regulatory actions and policy interventions to reduce food waste.

3 - Truth OR Dare? Managing Engagement and Verification Efforts Under Supplier Self-Reporting

Elif Kuscü, University of Zurich, Zurich, Switzerland, Tim Kraft, Tarkan Tan, Ozge Tuncel

To improve supply chain sustainability, buyers often rely on audits to assess performance and determine their engagement levels. However, traditional audits frequently fall short in uncovering the truth. Consequently, large firms such as health-technology provider Philips no longer use audits to ensure supplier compliance but instead require suppliers to self-report their sustainability performance. While novel, this practice has the potential to introduce new challenges, such as bias in suppliers' self-reports.

Motivated by this shift, we develop game-theoretic models to examine how a buyer should respond to a supplier's self-reported sustainability claims under information asymmetry. Specifically, we ask: How should buyers manage their efforts between engaging with suppliers and verifying their claims when reports may be strategically distorted? We analyze two settings: (1) a supplier self-reporting model in which the supplier may misrepresent their sustainability capability, and (2) a buyer verification model in which the buyer can adjust engagement based on supplier characteristics and invest in assessing the supplier's true capability.

Our analysis shows that supplier misreporting can induce higher buyer engagement and, counterintuitively, lead to improved sustainability outcomes. In some cases, the buyer achieves better results under asymmetric information than full transparency. These findings challenge the view that accurate reporting is an end in itself and highlight its role as a mechanism for driving improvement. From a managerial perspective, our findings illustrate the value of adaptive engagement strategies. Rather than focusing solely on verifying reports, buyers may benefit from a more nuanced approach that balances information gathering with proactive support.

4 - A Dynamic Multi-Agent Framework for Collective ESG Investment: Mitigation, Greenwashing, and Strategic Trade-Offs

Pengyu Wang, University of Washington, Foster School of Business, Seattle, WA, United States, Shi Chen

This paper develops a dynamic multi-agent decision framework to study collective ESG (Environmental, Social, and Governance) investment strategies in a setting where each firm's actions influence, and are influenced by, the broader population. We consider a large number of heterogeneous firms whose ESG-related decisions – ranging from genuine mitigation efforts to greenwashing – generate economic externalities such as climate risk and violations of sustainability standards in global supply chains. These externalities, in turn, affect the payoffs of all firms through their impact on an ESG-focused financial market. The systemic nature of ESG risks couples firms' decisions dynamically, making each firm's optimal strategy dependent on the collective behavior of others. Within this framework, we examine how firms' capital and ESG investment behaviors co-evolve over time. We highlight and analyze three key trade-offs: (1) the intertemporal tension between short-term costs and long-term ESG benefits; (2) the strategic balance between authentic mitigation and reputational greenwashing; and (3) the free-rider problem that emerges from shared mitigation efforts. Our findings provide insights into the design of effective ESG incentives that promote responsible investment and sustainable operations across global supply chains.

SB53

Building B Level 3 B313

Experimentation in Online Platforms

Invited Session

Revenue Management

Chair: Yifan Wu, Northwestern University, Evanston, IL, United States

Co-Chair: Gabriel Weintraub, Stanford Graduate School of Business, Stanford, CA, United States

1 - Reducing Symbiosis Bias Through Better A/B Tests of Recommendation Algorithms

David Holtz, Columbia Business School, New York, NY, United States, Jennifer Brennan, Yahu Cong, Yiwei Yu, Lina Lin, Yajun Peng, Changping Meng, Ningren Han, Jean Pouget-Abadie

It is increasingly common in digital environments to use A/B tests to compare the performance of recommendation algorithms. However, such experiments often violate the stable unit treatment value assumption (SUTVA), particularly SUTVA's "no hidden treatments" assumption, due to the shared data between algorithms being compared. This results in a novel form of bias, which we term "symbiosis bias," where the performance of each algorithm is influenced by the training data generated by its competitor. In this paper, we investigate three experimental designs—cluster-randomized, data-diverted, and user-corpus co-diverted experiments—aimed at mitigating symbiosis bias. We present a theoretical model of symbiosis bias and simulate the impact of each design in dynamic recommendation environments. Our results show that while each design reduces symbiosis bias to some extent, they also introduce new challenges, such as reduced training data in data-diverted experiments. We further validate the existence of symbiosis bias using data from a large-scale A/B test conducted on a global recommender system, demonstrating that symbiosis bias affects treatment effect estimates in the field. Our findings provide actionable insights for researchers and practitioners seeking to design experiments that accurately capture algorithmic performance without bias in treatment effect estimates introduced by shared data.

2 - The Design and Analysis of Switchback and Panel Experiments

Iavor Bojinov, Harvard Business School, Somerville, MA, United States

Marketplaces often exhibit complex interference patterns that in the worst case, can not be directly modeled or even adequately described. In such settings, practitioners have opted to use switchback experiments whereby the whole market is randomly assigned to treatment or control conditions for a short period, some outcome is recorded, and the process is repeated for T time periods. Although these experiments alleviate the challenges posed by the population interference, they often suffer from temporal interference — or carryover effects — whereby past treatments affect current outcomes. In this talk, I will share recent advances in the design and analysis of switchback experiments and their generalization to multiple experimental units, known as panel experiments. The talk will highlight insights from several new research papers and offer practical guidance for applying these methods in real-world settings.

3 - On Experiment Design Under Network Interference

Yuwei Luo, Stanford University, Stanford, CA, United States, Mohsen Bayati, William Overman, Sadeh Shirani, Ruoxuan Xiong

Randomized experiments are a powerful methodology for data-driven evaluation of decisions or interventions. Yet, their validity may be undermined by network interference. This occurs when the treatment of one unit impacts not only its outcome but also that of connected units, biasing traditional treatment effect estimations. This talk discusses a new framework to accommodate complex and unknown network interference and produce practical estimators.

4 - Switchback Price Experiments with Forward-Looking Demand

Yifan Wu, Northwestern University, Evanston, IL, United States, Ramesh Johari, Vasilis Syrgkanis, Gabriel Weintraub

We consider a retailer running a switchback experiment for the price of a single product, with infinite supply. In each period, the seller chooses a price p from a set of predefined prices that consist of a reference price and a few discounted price levels. The goal is to estimate the demand gradient at the reference price point, with the goal of adjusting the reference price to improve revenue after the experiment. In our model, in each period, a unit mass of buyers arrives on the market, with values distributed based on a time-varying process. Crucially, buyers are forward-looking with a discounted utility and will choose not to purchase now if they expect to face a discounted price in the near future. We show that forward-looking demand introduces bias in naive estimators of the demand gradient, due to intertemporal interference. Furthermore, we prove that there is no estimator that uses data from price experiments with only two price points that can recover the correct demand gradient, even in the limit of an infinitely long experiment with an infinitesimal price discount. Moreover, we characterize the form of the bias of naive estimators. Finally, we show that with a simple three-price level experiment, the seller can remove the bias due to strategic forward-looking behavior and construct an estimator for the demand gradient that asymptotically recovers the truth.

SB54

Building B Level 2 B216

Data-Driven Research in Revenue Management

Invited Session

Revenue Management

Chair: Zhiqi Zhang, Washington University in St. Louis, St. Louis, MO, United States

1 - Optimal Bundles in Land Freight Auctions

Yuetong Xu, MIT, Cambridge, MA, United States, Retsef Levi

Land freight is a critical transportation mode of shipping goods in the U.S., with trucks moving roughly 72.6% of total freight weight. This corresponds to \$940.8 billion of gross freight revenue, equivalent to 80.7% of the nation's freight bill in 2022. US Postal Services and many other agencies and companies employ freight delivery auctions with the goal of obtaining competitive market prices and maximize profits.

This work develops an optimization model to find the optimal bundles in second price freight auctions. An innovative and tractable integer program is provably able, for practically realistic size auctions, to identify bundles that ensure truthful bidding and minimize the overall payoff of the auctioneer.

2 - Bias-Variance Tradeoff of Matching Prior to Difference-in-Differences when Parallel Trends is Violated

Mingxuan Ge, University of Michigan - Ross Business School, Ann Arbor, MI, United States, Dae Woong Ham

The Difference-in-Difference (DiD) estimator is a popular estimator built on the "parallel trends" assumption. To increase the plausibility of this assumption, a popular and natural idea is to match treated and control units prior to a difference-in-difference. The intuition is that after matching, the matched set of treated units and control are more likely to have the same post-treatment outcome absent treatment, i.e., the parallel trends. While this issue has been studied previously, the focus was solely on the bias of such a practice. This paper generalizes prior bias results under weaker assumptions and further explores the variance and mean squared error (MSE) implications of matching DiD estimators. To assess the tradeoffs of matching prior to DiD, we compare the variances and MSE of three estimators: 1) classical DiD (no matching), 2) DiD after matching on observed covariates, and 3) DiD after matching additionally on pre-treatment outcome(s) under a linear structural equation model with unobserved time-varying confounders, allowing a breakage in parallel trends. Unlike previous studies that show matching on observed covariates X always reduces bias, this work shows matching on observed covariates X is not always recommended when considering variance due to a sample size tradeoff. We characterize full bias-variance tradeoffs among all three estimators and further give guidelines for determining to what extent the reduction in MSE from matching is likely to outweigh the associated costs.

3 - Data-Driven B2b Pricing

Ningyuan Chen, Rotman School of Management, Toronto, ON, Canada, Timothy Chan, Craig Fernandes

In B2B pricing, the price is typically determined through initial quote and subsequent bargaining. As a result, the prices in the transaction data are confounded by the customer's willingness-to-pay and thus endogeneous. To find the optimal (quoted) price and bargaining strategy, the approach adopted in the classical pricing literature, which first fits a demand function under the assumption that historical prices are exogeneous, does not apply. We develop a framework to learn the distribution of customer's willingness-to-pay and bargaining power from the data under confounding and then identify the optimal prices. We use the Nash bargaining to model the price forming process in the data. We identify the conditions of the dataset that lead to identifiable parameters, and show that even without identifiability, the optimal price can still be determined. We compare the efficiency of several methods with limited data.

4 - Bundling with Doubly Sparse Interactions: Estimation and Assortment Optimization for Bundling Under the Multivariate Mnl

Heng Zhang, Arizona State University, Tempe, AZ, United States, Mengxin Selene Wang

We study consumer choice behavior in bundling contexts, where customers may simultaneously purchase products across multiple categories. To capture both within-category substitution and cross-category complementarity, we present a reformulated multivariate multinomial logit (MVMNL) model. This framework extends the classical MNL model by modeling basket-level utility with Gumbel-distributed random variables, allowing us to account for complex interactions among products across different categories. A key innovation in our approach is to reduce the computational burden of this high-dimensional model by predicting choices in one category conditional on selections in others. This reformulation dramatically simplifies the full joint likelihood, making it more tractable for both estimation and optimization. We study the statistical properties of our approach. Recognizing that not all cross-category relationships are equally important, we impose a sparsity structure on interaction terms. This assumption enables us to develop efficient and scalable methods for parameter estimation. Building on our estimation procedure, we demonstrate how the learned interaction patterns can be directly exploited to design algorithms for operational decision-making. In particular, we illustrate the model's utility in assortment optimization and pricing strategies for high-dimensional retail settings.

5 - Balancing Act: Designing Sustainable Incentive Schemes for User Engagement and Revenue Optimization in Online Platforms

Xinyu Xu, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Qinlu Hu, Ni Huang, Renyu Zhang

User-generated content (UGC) platforms frequently use monetary incentives to boost user engagement, yet their impact on platform revenue remains under-explored. Our study addresses this gap by exploring the dual-sided effects of monetary incentives on both users and platforms and introduces and evaluates a novel monetary incentive scheme: ad-based incentives, where users receive rewards for watching advertisements. Unlike traditional monetary incentives (e.g., posting rewards), ad-based incentives generate direct platform revenue. Through two large-scale field experiments on a short-video platform, we analyze the effects of removing ad-based incentives and combined monetary incentives (a hybrid of traditional and ad-based schemes). Results show that removing ad-based incentives does not impact user engagement but significantly reduces platform revenue. Conversely, removing combined incentives reduces engagement but increases revenue, revealing a trade-off between user activity and profitability. Furthermore, Heterogeneous treatment effect analysis indicates greater impacts on women and older users (aged 50+). These findings provide actionable insights for designing balanced user acquisition and retention strategies.

SB55

Building B Level 3 B314

Recommendations, Fairness and Human-AI Interaction

Invited Session

Revenue Management

Chair: Akshit Kumar, Yale University, Jersey City, NJ, United States

1 - The Feedback Loop of Statistical Discrimination

Jackie Baek, New York University, New York City, NY, United States, Ali Makhdoumi

We study a dynamic model of interactions between a firm and job applicants to identify mechanisms that drive long-term discrimination. In each round, the firm decides which applicants to hire, where the firm's ability to evaluate applicants is imperfect. Each applicant belongs to a group, and central to our model is the idea that the firm becomes better at evaluating applicants from groups in which they have hired from in the recent past. We establish the firm's initial evaluation ability for each group to be a critical factor in determining long-term outcomes. We show that there is a threshold for which if the firm's initial evaluation ability for the group is below the threshold, the group's hiring rate decreases over time and eventually goes to zero. If the group starts above the threshold, then the hiring rate stabilizes to a positive constant. Therefore, even when two groups are ex-ante identical, a marginal difference in the firm's initial evaluation ability can lead to persistent disparities that exacerbate over time through a feedback loop. Importantly, the dynamic nature of our model allows us to assess the long-term impact of interventions. Our analysis highlights the importance of interventions that meaningfully alter evaluation ability, and our results imply that drastic short-term interventions are more effective compared to milder long-term interventions. We also find that smaller groups face inherent disadvantages, requiring a higher initial evaluation ability to achieve a favorable long-term hiring outcome and experiencing lower hiring rates even when they do.

2 - Why Students Reject AI for Human Counselors in College Applications: A Field Experiment

Hemanshu Das, Yale University, New Haven, CT, United States, Sofoklis Goulas, Faidra Monachou

AI is increasingly used to guide high-stakes educational decisions, yet its effectiveness depends on whether people follow its advice. We present the first large-scale field experiment with adolescents in this context, conducted across 14 public high schools. We compare the adoption of identical college-application recommendations from human counselors versus an AI-based algorithm. Contrary to the common assumption that objective, data-based recommendations favor AI, we find that algorithm aversion intensifies when recommendations are based on objective criteria (e.g., grades, admission chances) and dissipates when criteria are more subjective. We find that student perceptions of the recommender's intent strongly drive this aversion, consistently across scenarios and statistical approaches; perceptions of alignment with personal goals, ability, and comprehension also play significant roles. The results further reveal substantial heterogeneity in recommendation adoption rates. We observe stronger aversion among female students, students from rural schools, lower-GPA students, and those with stronger prestige-seeking attitudes. Using an optimization approach, we demonstrate how a policymaker can navigate the heterogeneity in recommendation adoption rates to optimally prioritize the assignment of human versus AI-based algorithmic recommenders, under varying social priorities and limited capacity of human counselors. We find that a targeting policy relying on few readily available

student and school features can approximate the first-best, personalized targeting policy effectively. These insights underscore the importance of understanding student preferences and trust in an AI system's intent, and of adopting hybrid approaches that blend human guidance with AI tools to design effective recommendation systems.

3 - How Agentic AI Impacts Customer-Intensive Services

Xiaofei Li, Zhejiang University, Hangzhou, China, People's Republic of, Ming Hu, Quan Yuan, Han Zhu

Artificial intelligence (AI) technologies are increasingly employed in customer-intensive services to automate auxiliary tasks and enhance service efficiency. We investigate a service system in which the overall service rate depends on both auxiliary and professional service times. AI effectively reduces the time allocated to auxiliary tasks. Our findings reveal that marginal improvements in AI capability always enhance market coverage but, somewhat surprisingly, may reduce service quality and service price. Depending on the service type, service price may follow different patterns---either always increasing, initially decreasing, or initially increasing and then decreasing---before eventually increasing, as AI capability improves. Only at a sufficiently high AI capability can the service provider enhance both service quality and service price. In contrast to traditional systems without AI, AI assistance tends to lower service quality unless AI capability reaches a sufficiently high level. Furthermore, we uncover an inverted U-shaped pattern in the optimal system utilization as AI capability increases and reveal a paradox: while AI improves the efficiency of auxiliary tasks, it can actually increase the provider's workload. Our study shows that with assistive AI to improve the operational efficiency of auxiliary tasks, the service provider likely reduces service quality and service price, but with more customers served on the demand side while creating more hiring needs for labor on the supply side.

SB56

Building B Level 3 B315

Designing Markets: Algorithms, Mechanisms, Information, and Experiments

Invited Session

Auctions and Market Design

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

Co-Chair: Jinglong Zhao, Boston University, Boston, MA, United States

1 - Automated Experimental Design with Optimization from Historical Data Simulations

Ruoxuan Xiong, Emory University, Atlanta, GA, United States

We study the problem of designing randomized experiments to estimate the causal effects of new interventions. The design problem involves selecting a treatment assignment mechanism, including assignment probabilities, in order to optimize a user-defined objective (e.g., maximizing the precision of a causal-effect estimator). A key challenge is that the objective is typically a black-box function of the design, depending on unknown outcome-generating processes and causal effects. To tackle this challenge, we propose a novel automated experimental design (AutoED) approach. AutoED evaluates the objective value of a candidate design by simulating experiments based on stationary historical control data and prior distributions of causal effects. Then a gradient-free method is used to iteratively optimize the design. We rigorously analyze the convergence behavior of AutoED and show how it depends on temporal and cross-unit correlations in historical data as well as on implementation specifics. In synthetic experiments on three application domains—digital platforms, health, and energy—AutoED can reduce the estimation error of treatment effects by up to 25% compared to the state-of-the-art benchmark designs. Our results reveal new design insights to improve efficiency. In particular, for multi-unit crossover experiments with multiple interventions, we find that efficiency is improved by sequentially rolling out (a) the same intervention across units and (b) different interventions on the same unit.

2 - The Power of Two in Token Systems

Süleyman Kerimov, Rice University, Houston, TX, United States, Itai Ashlagi, Omer Tamuz, Geng Zhao

In economies without monetary transfers, token systems serve as an alternative to sustain cooperation, alleviate free riding, and increase efficiency. This paper studies whether a token-based economy can be effective in marketplaces with thin exogenous supply. We consider a marketplace in which at each time period one agent requests a service, one agent provides the service, and one token (artificial currency) is used to pay for service provision. The number of tokens each agent has represents the difference between the amount of service provisions and service requests by the agent. We are interested in the behavior of this economy when very few agents are available to provide the requested service. Since balancing the number of tokens across agents is key to sustain cooperation, the agent with the minimum amount of tokens is selected to provide service among the available agents.

When exactly one random agent is available to provide service, we show that the token distribution is unstable. However, already when just two random agents are available to provide service, the token distribution is stable, in the sense that agents' token balance is unlikely to deviate much from their initial endowment, and agents return to their initial endowment in finite expected time. Our results mirror the power of two choices paradigm in load balancing problems. Supported by numerical simulations using kidney exchange data, our findings suggest that token systems may generate efficient outcomes in kidney exchange marketplaces by sustaining cooperation between hospitals.

3 - Learning to Design Information

Ce Li, Boston University, Boston, MA, United States

Classical information design models (e.g., Bayesian persuasion and cheap talk) require players to have perfect knowledge of the prior distribution of the state of the world. Our paper studies repeated persuasion problems in which the information designer does not know the prior. The information designer learns to design signaling schemes from repeated interactions with the receiver. We design learning algorithms for the information designer to achieve no regret compared to using the optimal signaling scheme with known prior, under two models of the receiver's decision-making. (1) The first model assumes that the receiver knows the prior and can perform posterior update and best respond to signals. In this model, we design a learning algorithm for the information designer with $O(\log T)$ regret in the general case, and another algorithm with $\Theta(\log \log T)$ regret in the case where the receiver has only two actions. (2) The second model assumes that the receiver does not know the prior and employs a no-regret learning algorithm to take actions. We show that the information designer can achieve regret $O(\sqrt{r\text{Reg}(T)T})$, where $r\text{Reg}(T)=o(T)$ is an upper bound on the receiver's learning regret. Our work thus provides a learning foundation for the problem of information design with unknown prior.

4 - Priority Matching in Kidney Exchange Markets

Elaine Liu, Stanford University, Stanford, CA, United States, Itai Ashlagi, Pengyu Qian

We study the kidney exchange market with a priority structure. In this market, one donor-patient pair arrives per period. The pair is type “willing” with probability $\$q$ and “unwilling” otherwise. Donors of “willing” pairs can donate prior to the patient receiving, and receive priority in return once they donate; “unwilling” pairs are only willing to donate at the time or after the patient receives a kidney. Compatibility probability is $\$p$ between any patient and donor, except the pair that arrives together.

In this work, we carefully define “priority” by proposing three different matching algorithms: weak priority, strict priority, and strictest priority. The weak priority algorithm greedily maximizes for the most donations, as long as it does not harm priority patients. The strict priority algorithm greedily finds the fastest way to help priority patients. The greedy nature makes these two algorithms computationally efficient. On the other hand, strictest priority finds the absolute fastest way to help priority patients by searching over the entire market. We analyze the expected waiting time for willing and unwilling patients under all three algorithms. We conclude that, as long as the priority is not overly aggressive, the priority market structure benefits both types of patients alike. Our results have broader implications in self-serving markets beyond kidney exchange.

SB57

Building B Level 3 B316

Technology's Impact on Social Structures and Organizational Dynamics

Invited Session

Social Media Analytics

Chair: Reihane Boghrati, Arizona State University, Tempe, AZ, United States

1 - Beyond Productivity: Design and Effects of Algorithmic Output on Cognitive Engagement and Distinctiveness

Anuschka Schmitt, The London School of Economics and Political Science, London, United Kingdom, Thiemo Wambsganss, Andreas Janson, Krzysztof Z. Gajos

Can GenAI help humans without telling them what to do? Prevalent forms of algorithmic output provide a user with 'one best' solution that can induce the user to fixate on the algorithmic output and neglect critical reasoning. Cognitive engagement is important, though, in consideration of long-term outcomes such as individual learning and intellectual diversity. As part of a series of field experiments in the context of a business pitch writing task, we investigate how alternative designs of algorithmic output impact three interrelated outcomes: task performance, cognitive engagement, and distinctiveness of the written pitches. We explore the potential of (i) reciprocal algorithmic output providing feedback in form of open-ended questions and pro and contra arguments compared to (ii) a 'default' design of solution-focused algorithmic output improving users' writing without any further advice. We show that individuals can improve their writing regardless of the type of algorithmic output exposed to. However, individuals exposed to reciprocal algorithmic output engage significantly more with the task and the algorithmic output and write more distinct business pitches compared to individuals exposed to contrastive algorithmic output. Together, these results indicate that there are different paths that enable productivity gains through GenAI. Next to individual-level upsides, reciprocal algorithmic output can have significant long-term and group-level benefits including the preservation of intellectual distinctiveness. Particularly relevant to educational and organizational settings where cognitive engagement is paramount to learning and skill development, our study illustrates how algorithmic output may serve as a critical stimulator rather than attenuating human agency.

2 - Elements of Persuasiveness in Kickstarter Pitch Videos: a Multi-Modal and LLM-Based Approach

Maryam Rahmani Moghaddam, University of Delaware, NEWARK, DE, United States, Weiguo (Patrick) Fan

This study explores persuasive strategies in crowdfunding pitch videos using a multi-modal framework that incorporates visual and auditory analysis. Guided by the AIDA (Attention, Interest, Desire, Action) model, we focus on identifying persuasive elements such as product demonstrations, statistical claims, and credibility cues using large language models (LLMs) like GPT-4. Videos are segmented into scenes using the Google Vision API, and a representative frame from each scene is analyzed for persuasive content. Acoustic features such as pitch, intensity, and formants are extracted to capture delivery style, while visual features include human presence, colorfulness, and brightness, among others. Our aim is to develop an interpretable model that integrates these features to predict campaign success and provide insights into persuasive communication in video-based crowdfunding.

3 - The Impact of Platform-Driven Movements on Gender Leadership Dynamics

Reihane Boghrati, Arizona State University, Tempe, AZ, United States

Digital information systems, particularly social media platforms, can turn individual stories into widespread movements that drives real change within organizations. Drawing on social movement theory in the context of digital communication, we examine how the MeToo movement, a social media-driven movement, influenced female leadership in the workplace. Using a quasi-experimental approach, we compile and analyze a comprehensive dataset of individual career paths and firm characteristics. Results show an increase in female representation in leadership positions following the MeToo movement. This effect is moderated by factors such as individual industry backgrounds, career mobility, and industry gender composition at leadership levels. Post-MeToo, women saw a 40% increase in the likelihood of attaining leadership roles. Moving from non-IT roles into IT roles (27% increase) as well as moving from non-IT industry to IT industry (11% increase) benefited women post-MeToo. The MeToo movement had a more pronounced positive effect on women's career trajectories in organizations that already had a strong female presence in leadership (64% increase post-MeToo). Our findings provide valuable insights into how the MeToo, a social media-driven movement influenced, as a secondary impact, workplace gender dynamics.

4 - Teaming and Competing with AI: Communication Design in Hybrid Multi-Agent Environments

Sungeun (Selena) Han, Arizona State University, Tempe, AZ, United States, Tian Lu, Reihane Boghrati

As generative-AI (Gen-AI) becomes increasingly integrated into organizational workflows, understanding how AI agents communicate with human teammates has become a critical research frontier. Prior studies have primarily focused on dyadic (one-human, one-agent) collaborations, where emotional and logical communication styles have been shown to shape trust, alignment, and performance. However, many emerging applications involve hybrid multi-agent teams, where a human collaborates with one AI agent while simultaneously competing with others. In these cooperative settings, communication dynamics are more complex, and established findings from dyadic contexts may not generalize. This study investigates how Gen-AI's emotional and logical communication styles shape human perceptions and collaborative outcomes in a multi-agent hybrid cooperation game. We conduct a controlled lab experiment using the social deduction game Werewolves, a high-uncertainty, high-interdependence environment that requires strategic coordination and competition. By systematically varying the AI agent's communication style, we aim to examine its impact on trust, perceived competence, satisfaction, collaborative behavior, and team performance. Emotional communication enhances social presence but may trigger strategic distrust; logical communication improves decision alignment but may limit social cohesion. These dynamics challenge established design principles for AI communication in dyadic collaboration. We contribute to the literatures on human-AI teaming, communication design, and collective intelligence by showing that communication strategies must adapt to the structural complexity of multi-agent environments to support effective and resilient teamwork.

SB58

Building B Level 3 B317

Advances in Scheduling

Contributed Session

Contributed

Chair: Renjie Yu, The University of Sydney, Sydney, Australia

1 - Post-Earthquake Restoration Problem of Water Distribution Network

Peihong Xiao, National University of Singapore, Singapore, Singapore

Restoring water distribution systems is a critical yet complex task in post-disaster recovery. Timely pipe restoration is essential to ensure clean water access, supporting both public health and disaster relief. However, challenges such as limited resources, uncertain pipe damage, and urgent timelines complicate the process.

This study addresses the water distribution network restoration problem under uncertainty, where pipe damage is initially unknown and progressively revealed via sensors at demand nodes. Since delays in repairing damaged pipes can disrupt downstream supply and cause significant losses, the goal is to determine an optimal restoration sequence that minimizes the total weighted delay.

We first model this as a partially observable Markov decision process (POMDP). However, solving large-scale POMDPs is computationally infeasible. To address this, we propose a two-layer open-loop policy that uses current information without relying on future observations. In the outer layer, a Bayesian network updates the belief state of each pipe based on the latest sensor data. This updated state is then passed to the inner layer, a scheduling model that computes the pipe restoration sequence assuming no future observation updates. As new observation becomes available, the process iterates, dynamically adapting decisions while maintaining tractability. We also identify structural properties of the inner-layer scheduling problem and develop a heuristic to efficiently generate restoration sequences.

Numerical experiments on randomly generated networks with varying topologies and information settings show that our approach significantly outperforms benchmark strategies in both efficiency and total cost.

2 - Port Terminal Appointment Scheduling Problem with Random Arrival and Processing TIMES

Mustafa Gurbuz, Zaragoza Logistics Center, ZARAGOZA, Spain, Lygia Brønneberg Velez

Motivated by the resolution published by the Brazilian National Agency of Oil in July of 2022, we study the Port Terminal Appointment Scheduling Problem (PTASP) where ports strive for maximizing profits while offering clients such as ship owners quality service (e.g., appointments that meet clients' requirements, reduced waiting times).

Our model considers clients requests, contractual agreements, delays in arrivals and processing times distributions, and pre-defined berthing sequence rules as input parameters. We assume that the processing times for activities such as loading and unloading at the port are stochastic. Different to most existing work in the field, the "actual" ship arrival times are also stochastic, defined by random deviations from the "scheduled" arrival time in our model.

In order to reduce solution time, we propose heuristics employing a "Cluster First, Schedule Second" methodology and use simulation optimization to perform numerical analysis using real data provided by a major oil and gas transportation company in Brazil. The decision variables are the binary decisions to accept or reject requests from clients and scheduled dates of arrival if accepted.

Our results show that information sharing (e.g., volume information) between the client and the terminal, clients flexibility in the appointment time window, coordination between scheduling and operational teams, and static/dynamic assignment of appointments may have significant impact on the financial and operational performance of the terminal. Interestingly, increased variability in ship arrival times might lead to improvements in profits in some cases.

3 - Appointment Scheduling for Logistics Parks Under Truck Heterogeneity and Order Fulfillment Delay

Yujiao Sun, Guangdong University of Foreign Studies, Guangzhou, China, People's Republic of

Appointment scheduling is critical for logistic parks of manufacturing industry firms to manage inventory turnover and schedule service objects. This work studies the appointment scheduling problem of a logistic park considering truck heterogeneity and order fulfillment delay, where the number of time windows in a day, the number of each type of appointment slot, and the service schedule in each time window are decided. The objective is to minimize the expected total cost of the truck waiting time, inventory, and overtime. We develop a shortest service time first scheduling strategy to determine the service sequence in each time window and prove that it generates optimal scheduling solutions. Based on our proposed scheduling strategy, we develop a recursive calculation to formulate the cost function, which provides a tractable approach to solving combinatorially complex problems. The proposed model is solved by a dual-selection based genetic algorithm (DSGA). The performance of the DSGA is validated by comparing it to a genetic algorithm using a single tournament selection operation. Our experimental results show that the DSGA enlarges the size of subgroups with better fitness and therefore performs satisfactorily.

4 - The Car Resequencing Problem with Makespan Minimization

Xinyi Guo, Tsinghua University, Beijing, China, People's Republic of, Jean-François Côté

In the physical car resequencing problem, a sequence of cars is readjusted to a new one via a buffer placed between two adjacent production shops to minimize total cost. This paper introduces a new version with a minimum makespan objective, which arises when carmakers adopt a buffer with multiple first-in-first-out lanes and a reverse lane to rearrange cars to a given sequence. The reverse lane enhances resequencing flexibility by permitting car re-entry in the buffer but increases computational intractability. We propose a time-space network formulation and valid lower bounds for this problem, investigate its computational complexity, as well as develop an exact Branch-and-Price-and-Cut approach integrated with Constraint Programming (BPC-CP). In the BPC-CP approach, a Constraint Programming model is solved to find a feasible solution using the partial integer information from the Branch-and-Price to mitigate its over-fractional issue. Otherwise, valid inequalities are generated and added when the partial information is verified to be infeasible. Computational experiments demonstrate the effectiveness of the proposed approach.

5 - Drones for Precision Agriculture

Srinivasa Prasanna, IIIT-Bangalore, Bengaluru, India

We discuss the special requirements of drone scheduling for precision agriculture. There are requirements of multiple take-offs/landings at uncertain locations, depending on local crop health, even at the level of each individual plant. The wide area coverage requirements result in unique challenges, different from standard scheduling problems. Multiple heterogeneous drones have to be scheduled in a time and energy optimal fashion, covering the field, with uncertain load -cameras, treatments chemicals, etc. Each drone has to be scheduled with close proximity to 10+ others, for adequate speed in area coverage. This, coupled with return to hubs for charging, poses a challenging optimization problem, with 1000's of variables in large farmlands, to be covered in a couple of hours or less. Results from several approaches will be presented in the talk

6 - Single Machine Scheduling with Fixed Energy Recharging Times to Minimize the Number of Late Jobs and the Number of Just-in-Time Jobs

Renjie Yu, The University of Sydney, Sydney, Australia, Daniel Oron

We study single-machine scheduling problems where processing each job requires both processing time and rechargeable energy. Subject to a predefined energy capacity, energy can be recharged after each job during a fixed recharging period. Our focus is on two due date-related scheduling criteria: minimizing the number of late jobs and maximizing the weighted number of jobs completed exactly at their due dates. This study aims to analyze the parameterized tractability of the two problems and develop fixed-parameter algorithms with respect to three natural parameters: the number of different due dates $\nu(d)$, the number of different processing times $\nu(p)$, and the number of different energy consumptions $\nu(e)$. Following the proofs of $\mathcal{NP}$ -hardness across several contexts, we demonstrate that both problems remain intractable when parameterized by $\nu(d)$ and $\nu(p)$. To complement our results, we show that both problems become fixed-parameter tractable (FPT) when parameterized by $\nu(e)$ and $\nu(d)$, and are solvable in polynomial time when both $\nu(e)$ and $\nu(p)$ are constant.

SB59

Building B Level 3 B318

Content and Recommendation Systems

Contributed Session

Contributed

Chair: Rama Samantapudi, 320 Autumn Oak Way, 78628

1 - Context-Aware Recommendation System for Korean Media Content in Language Education: a Case Study on the Kokiri App

Seungye Bae, Koylabs, Seoul, Korea, Republic of

Korean popular media such as dramas, variety shows, and idol videos are increasingly used by global learners as immersive tools for studying the Korean language. However, selecting suitable content that aligns with each learner's language level, interests, and learning context remains a challenge. To address this, we developed a context-aware recommendation system for language education, designed to personalize the delivery of Korean media content in the KOKIRI app—an AI-powered platform for Korean language learning.

The system integrates user behavior data, such as viewing history and learning progress, with content metadata, including difficulty level, topics, usage scenarios, and speaking styles. Using a hybrid approach that combines content-based filtering with contextual features, the system recommends short-form K-content that best matches each learner's needs and motivations.

This recommendation engine was deployed in a real-world environment and evaluated with active KOKIRI users. The results showed significant improvements in learner engagement, content consumption continuity, and satisfaction with the learning experience. In particular, context-aware recommendations led to longer viewing times and higher rates of voluntary content bookmarking.

This study contributes a scalable framework for personalized language education using cultural media, especially for non-native learners. It demonstrates how the integration of machine learning and cultural context can enhance user experience in EdTech platforms. We present our system architecture, data pipeline, and field results to share insights for both academic and practical applications.

2 - The Impact of Recommendation Systems on Content Consumption Patterns in Serialized Webtoon Content

Bongjin Sohn, Korea University, Seoul, Korea, Republic of, Antino Kim, Gunwoong Lee

This research examines the impact of recommendation systems on user consumption behavior in the context of serialized content platforms. Unlike streaming (or short-form), which allows for continuous consumption, serialized formats such as digital comics (i.e., webtoons) impose a nested temporal structure characterized by mandatory waiting periods and recurring short-session consumptions. These temporal constraints compel users to make frequent consumption decisions under a limited budget and time, increasing cognitive fatigue and altering content engagement patterns. Leveraging a field experiment on a leading webtoon platform in Korea, we compare recommendation-based versus popularity-based content displays to assess their effects on user conversion behavior (i.e., paid or alternative content consumption). The serialized release structure, particularly the distinction between initial publication and subsequent availability, enables us to examine how the timing of algorithmic exposure moderates the effectiveness of recommendations. Preliminary findings suggest that recommendation systems enhance conversion by either reinforcing early-stage engagement or facilitating content switching as decision fatigue accumulates. These results highlight that recommendation systems serve not only as content-matching tools but also as adaptive behavioral interventions in environments requiring repeated decision-making, suggesting important implications for platform operators in designing recommendation strategies aligned with the temporal dynamics of serialized content.

3 - The Ripple Effect of Live Commerce: Market-Wide Impacts of Broadcasting Activity in E-Retail

Mingi Song, Korea University Business School, Seoul, Korea, Republic of, Miaoze Han, Gunwoong Lee, Keongtae Kim

Live Streaming Commerce (LSC) has emerged as a transformative force in digital retail, yet its systemic effects on market dynamics and competitive landscapes remain underexplored. This research examines how LSC broadcasting activity impacts seller performance, competitive interactions, and platform-wide outcomes using comprehensive longitudinal data from a major South Korean e-commerce marketplace. Through rigorous causal inference analysis, we reveal that LSC broadcasting substantially enhances participating sellers' performance by capturing increased consumer attention. These benefits exhibit pronounced heterogeneity—smaller, less-established stores experience disproportionately larger gains compared to dominant sellers. We also identify asymmetric spillover effects: LSC broadcasting creates positive externalities for large competing sellers while generating negative impacts on smaller non-participating rivals. Despite these heterogeneous spillovers, LSC activity produces a net positive effect on overall platform performance, primarily driven by substantial gains among broadcasting sellers, especially smaller ones. This study advances LSC research by quantifying cross-seller and ecosystem-wide effects, providing empirical evidence for attention-based mechanisms and their market-position-dependent value, and demonstrating LSC's role as a market-restructuring innovation that reshapes competitive dynamics. Our findings position LSC broadcasting as a potentially democratizing mechanism in e-commerce that can level competitive playing fields.

4 - The Interplay of The Demand Effects of Recommendations and Advertising

Panagiotis Adamopoulos, Emory University, Atlanta, GA, United States, Vilma Todri

Recommender systems and advertising are independently considered among the most successful persuasion techniques frequently employed by platforms and marketers, owing to their significant influence on consumer behavior. Yet, our scientific understanding of the interplay of these business tactics remains rather limited. In this study, we investigate the relationship between advertising and the effectiveness of recommendations addressing an important gap in the literature. Specifically, we examine whether advertising expenditures relate to the effectiveness of system-generated recommendations towards increasing demand. Interestingly, our results show that even though both recommendations and advertising independently have a positive impact on demand, combining these different persuasion techniques might lead to advertising reducing the effectiveness of recommendations. However, advertising in different channels, earned media, and marketing promotions that increase the transaction utility of consumers might not cannibalize the positive effect of recommender systems on product demand. This discovered interplay of the advertising and recommender systems demand effects is likely due to the consumers' elevated levels of persuasion knowledge when exposed to multiple same-channel consumer persuasion techniques. The findings are supported by various identification strategies and robustness checks. We further validate the findings and underlying mechanisms in a lab experiment. Our study enhances the current understanding of the demand effects of recommendations and extends the persuasion knowledge theory to recommender systems, while having important implications for platforms and marketers and suggesting new directions for future research.

5 - Scalable and Modular Guided Search Query Recommendations via Collaborative User Navigational Behavior Graph Modeling and Modular Signal Merging

Rama Samantapudi, Walmart, Georgetown, TX, United States

This paper presents a scalable framework for constructing guided search query recommendations by leveraging users' real-world search navigational behavior. The proposed system captures collaborative query sequences from millions of user sessions, utilizing them to generate candidate queries through sequence modeling based embedding similarity retrieval techniques. To enhance relevance and diversity, random walks are performed over a directed graph constructed from observed query transitions, effectively simulating synthetic query sequences that preserve semantic coherence. These sequences are then encoded via a Skip-Gram word2vec model to produce dense, contextualized query embeddings.

The framework also integrates signal merging to incorporate global and local ranking using query popularity, conversion rates, and regional engagement scores to surface the most relevant query suggestions. A personalization layer is added which further refines guided search suggestions by leveraging individual user-profiles, historical preferences, and recent activity, ensuring tailored and meaningful query navigation. This multi-layered approach aims to improve the relevance, contextual alignment, and personalization of search query experiences in a modular way enabling us to independently develop each component through customized training schedules and tailored inference pipelines thereby optimizing and reducing overall system costs when compared to existing systems.

Simulation studies on e-commerce data demonstrates that the proposed system can significantly improve relevance, personalization, and user engagement. This work highlights the effectiveness of sequence modeling on graph-based collaborative query walks for injecting users' navigational search journey signals, and the cost-optimized modular approach which supports adding several global, regional & personalized signals to surface contextual, popular and personalized search experience.

SB60

Building B Level 3 B319

AI in Societal Systems

Contributed Session

Contributed

Chair: Elia Rezaeian, Stevens Institute of Technology, 5609 Jefferson St, Hoboken, NJ, 07093, United States

1 - Bridging the Digital Divide: the Role of AI Literacy and Trust in Voice Assistant Adoption

Hsin-yi Tsai, National Cheng Kung University, Tainan, Taiwan, Chien-Wen Yuan, po-yu Hsieh, An-Yu Liu, Chieh-Ren Hsiao, Yu-Ting Chen

AI (Artificial Intelligence) technologies have become integral to daily life, yet few studies explore the role of AI literacy in technology adoption. Existing research on AI literacy primarily focuses on technical skills, overlooking essential aspects like ethics, privacy, and risk awareness. This gap is critical as AI systems, such as voice assistants, increasingly process personal data, highlighting the need for safe and responsible usage. Addressing these concerns through AI literacy can promote trust and ethical awareness, supporting smoother adoption.

Voice assistants use speech recognition and AI technologies to improve convenience and quality of life. However, their growing popularity also exposes a digital divide, where certain populations lack the necessary skills or trust to fully engage with these technologies. This divide may lead to inequitable access, emphasizing the need for inclusive digital literacy programs that extend beyond technical proficiency to include ethics and privacy awareness.

This study examines how AI literacy and perceived trust influence the intention to use AI voice assistants. Integrating AI literacy into the Unified Theory of Acceptance and Use of Technology (UTAUT), an online survey in Taiwan ($N = 1,173$) reveals that AI literacy positively impacts perceived trust and effort expectancy, which in turn enhance performance expectancy. Furthermore, AI literacy, perceived trust, and performance expectancy significantly predict usage intentions. These findings underscore the importance of AI literacy for building trust, reducing risk perceptions, and bridging the digital divide, leading to more equitable and responsible technology adoption. Detailed suggestions for promoting the use of voice assistants are provided.

2 - Fraud-Proof Revenue Division on Subscription Platforms

Nicholas Teh, University of Oxford, Oxford, United Kingdom, Abheek Ghosh, Tzeh Yuan Neoh, Giannis Tyrovolas

We study a model of subscription-based platforms where users pay a fixed fee for unlimited access to content, and creators receive a share of the revenue. Existing approaches to detecting fraud predominantly rely on machine learning methods, engaging in an ongoing arms race with bad actors. We explore revenue division mechanisms that inherently disincentivize manipulation. We formalize three types of manipulation-resistance axioms and examine which existing rules satisfy these. We show that a mechanism widely used by streaming platforms, not only fails to prevent fraud, but also makes detecting manipulation computationally intractable. We also introduce a novel rule, ScaledUserProp, that satisfies all three manipulation-resistance axioms. Finally, experiments with both real-world and synthetic streaming data support ScaledUserProp as a fairer alternative compared to existing rules.

3 - Balancing Transparency and Efficiency: the Role of Explainable AI in Multi-Sided Logistics Platforms

Wei Ziyi, Southeast University, Nanjing, China, People's Republic of

With the growing integration of Artificial Intelligence (AI) in modern logistics platforms, AI-driven algorithms are increasingly responsible for core operational decisions such as order allocation, route optimization, and merchant recommendations. While these systems enhance operational efficiency, their opaque “black-box” nature often undermines user trust, raising concerns among merchants and consumers regarding fairness and decision legitimacy. Explainable AI (XAI) emerges as a potential solution by providing transparent insights into algorithmic decisions. However, existing literature predominantly focuses on duopoly market structures, overlooking the complex multi-sided interactions inherent in platform ecosystems.

This study extends the analysis of XAI to a triadic platform environment involving the platform, merchants, and consumers. We develop a multi-stage game-theoretic model where the platform strategically chooses its level of XAI transparency. The model evaluates how different transparency strategies influence stakeholder behaviors, trust, and overall welfare. Furthermore, we assess the implications of regulatory interventions aimed at mandating XAI levels. Our findings suggest that full transparency may not always be socially optimal; instead, a flexible regulatory framework that accommodates platform heterogeneity and interaction dynamics may yield superior outcomes. The study contributes to both the theoretical understanding of XAI in platform economies and the practical design of governance mechanisms in algorithmic decision-making environments.

4 - When Transparency Talks: Explainability in Human-AI Hybrid Decision-Making on Crowdfunding Platforms

Saunak Basu, University of Nevada Las Vegas (UNLV), Las Vegas, NV, United States, Ashmita Biswas

The increasing integration of artificial intelligence (AI) into digital platforms has transformed how users evaluate complex decisions, especially in high-stakes environments like healthcare crowdfunding. While AI can enhance predictive accuracy, its black-box nature can undermine trust and user agency. This study examines whether *explainability* in AI recommendations fosters more effective human-AI collaboration, influencing donor decision-making and campaign success on medical crowdfunding platforms.

Using a unique field dataset from a global healthcare crowdfunding site, coupled with a natural experiment involving the staged rollout of explainable AI (XAI) modules, we employ a difference-in-differences identification strategy to measure causal effects on donor engagement, campaign funding velocity, and donor retention. We further test heterogeneity by sentiment in campaign imagery and narrative language complexity, linking to prior work on visual persuasion.

Our findings provide robust empirical evidence that XAI boosts donor trust and funding participation, particularly in campaigns with emotionally ambiguous narratives. Contrary to expectations, the effect is *stronger* in low-literacy regions, suggesting that transparency—not just accuracy—can democratize digital philanthropy.

We highlight our contribution by foregrounding the behavioral impacts of AI explainability in hybrid decision settings and by highlighting its potential to reduce inequities in digital access. Our work offers actionable insights for platform designers seeking to optimize human-AI interfaces for transparency, fairness, and engagement.

5 - AI and the Extended Workday: Productivity, Contracting Efficiency, and Distribution of Rents

Junyoung Park, Auburn University, Auburn, AL, United States, Wei Jiang, Rachel Xiao, Shen Zhang

This study investigates how occupational AI exposure impacts employment at the intensive margin, i.e., the length of workdays and the allocation of time between work and leisure. Drawing on individual-level time diary data from 2004–2023, we find that higher AI exposure—whether stemming from the ChatGPT shock or broader AI evolution—is associated with longer work hours and reduced leisure time, primarily due to AI complementing human labor rather than replacing it. This effect is particularly pronounced in contexts where AI significantly enhances marginal productivity and monitoring efficiency. It is further amplified in competitive labor and product markets, where workers have limited bargaining power to retain the benefits of productivity gains, which are often captured by consumers or firms instead. The findings question the expectation that technological advancements alleviate human labor burdens, revealing instead a paradox where such progresses compromise work-life balance.

6 - Congestion Games with Dueling Preference Feedback: Equilibrium, Learning, and Provable Efficiency

Yunian Pan, New York University, BROOKLYN, NY, United States, Quanyan Zhu

We propose a new class of congestion games driven by pairwise preference feedback rather than scalar cost functions. Motivated by limitations of classical Wardrop models in capturing complex user behavior, we define the *Borda Wardrop Equilibrium* (BWE) as the equilibrium condition where traffic flows align with Borda-aggregated ordinal preferences. We show that BWE arises as the minimizer of a convex *Borda-Beckmann potential*, generalizing classical potential function arguments for congestion games. We further design an online learning algorithm based on dueling feedback, which achieves $\mathcal{O}(T^{2/3})$ regret and provable convergence to BWE in the large-population limit. Our results establish structural properties of preference-driven congestion games, extend potential game theory to ordinal settings, and demonstrate efficient learning under minimal informational assumptions. This work provides a mathematical foundation for studying non-cost-based congestion phenomena under human-in-the-loop dynamics.

7 - From Black Box to Transparency: Experimental Evidence on the Role of AI Explainability in Shaping Clinicians Perceptions

Olya Rezaeian, Stevens Institute of Technology, Hoboken, NJ, United States, Onur Asan, A. Emrah Bayrak

In recent years, explainability has emerged as a critical factor influencing the adoption and effectiveness of AI-based Clinical Decision Support Systems (CDSS) in healthcare. This mixed-methods study investigates how varying levels of AI explainability affect clinicians' trust, diagnostic performance, and perception in the context of

breast cancer diagnosis. We conducted a human-subject experiment using a custom-built web-based AI system that delivered diagnostic suggestions. Quantitative data were collected from 28 clinicians. To further explore clinicians' interpretations and expectations, semi-structured interviews were conducted and analyzed using thematic analysis. The findings indicate that increasing levels of explanation do not consistently improve clinicians' trust or diagnostic accuracy. High AI confidence scores substantially increased trust but also led to overreliance, ultimately reducing diagnostic performance. In contrast, low confidence scores decreased trust and agreement while increasing the time spent on decisions, reflecting more cautious behavior. Certain explainability features were linked to higher cognitive load and stress. Additionally, demographic factors such as age, gender, and professional role influenced how participants perceived and interacted with the system. Notably, some clinicians exhibited trust in the AI even when it was incorrect, highlighting the complex and sometimes counterintuitive effects of explainability. Qualitative findings revealed clinicians' desire for explanations that are transparent, actionable, and aligned with their clinical reasoning. These results emphasize the need for explainable AI systems that not only inform but also support human decision-making in contextually meaningful ways.

SB61

Building B Level 4 B401

QSR Best Refereed Paper Competition

Award Session

Quality, Statistics and Reliability

Chair: Jia Liu, University of Florida, Gainesville, FL, United States

Co-Chair: Shancong Mou, University of Minnesota Twin Cities, Minneapolis, MN, 55455, United States

Co-Chair: Xiaochen Xian, Georgia Institute of Technology, Atlanta, GA, United States

1 - CrowdLLM: a Synthetic Crowd Simulator for Crowdsourcing with LLM Workers Augmented with Lightweight Generative Models

Feng (Ryan) LIN, University of Washington, Seattle, WA, United States, Hanming Zheng, Keyu Tian, Congjing Zhang, Li Zeng, Shuai Huang

LLMs have sparked extensive interest to create synthetic crowds in human decision-making tasks. Using crowdsourcing as a testbed, we propose the CrowdLLM framework that can simultaneously enhance the accuracy and diversity of the collective decision-making outcomes by leveraging pretrained LLMs and generative models.

2 - Scalable First-Order Method for Certifying Optimal K-Sparse GLMs

Jiachang Liu, Cornell University, Ithaca, NY, United States, Soroosh Shafiee, Andrea Lodi

This paper addresses certifying optimality for sparse GLMs using a novel first-order proximal gradient algorithm. Our method solves the perspective relaxation within a branch-and-bound framework, offering linear convergence and GPU compatibility. It significantly accelerates dual bound computations and provides effective optimality certificates for large-scale problems, enhancing model interpretability and trustworthiness.

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

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

We propose a domain-informed reinforcement learning framework for preventive maintenance and inspection planning in multi-component systems with stochastic degradation and economic dependence. By integrating hierarchical policies, monotonicity constraints, and Bayesian degradation models, our method addresses vast continuous state-action space challenges. Numerical and case studies validate substantial outperformance over existing methods.

4 - Physics-Augmented Multi-Task Gaussian Process for Modeling Spatiotemporal Dynamics

Xizhuo Zhang, The University of Tennessee, Knoxville, TN, United States, Bing Yao

Modeling high-dimensional spatiotemporal data is challenging due to complex geometries and variable interdependence. We propose a physics-augmented multi-task Gaussian Process (P-M-GP) that integrates geometry-aware modeling, inter-task learning, and physics-based regularization. P-M-GP achieves superior performance over data-driven approaches when applied to cardiac electrodynamic modeling.

SB62

Building B Level 4 B402

Scientific Domain-Informed Machine Learning for Complex Systems

Invited Session

Quality, Statistics and Reliability

Chair: Minhee Kim, University of Florida, Gainesville, FL, United States

1 - Dynamic Calibration of Digital Twin via Stochastic Simulation: a Wind Energy Case Study

Yongseok Jeon, North Carolina State University, Raleigh, NC, United States, Sara Shashaani, Eunshin Byon, Pranav Jain

This study presents an approach to dynamically calibrating a digital twin to support decision-making in systems operating under uncertainty. The framework integrates uncertainty by turning a physics-based model into a stochastic simulation, where independent variables that represent the environmental conditions may be nonstationary whereas target variables are conditionally stationary. The calibration task itself is formulated as a simulation optimization problem that we solve using a root-finding strategy. As a case study, we apply the suggested framework to the prediction of short-term power deficit, known as the wake effect, in wind farms using real-world data and demonstrate the robustness of the proposed framework. Besides advancing digital twin research, the presented methodology contributes to wind farm wake steering strategy by enabling accurate short-term wake effect prediction.

2 - Bll-Pinn: Bayesian Last-Layer Physics-Informed Neural Networks for Uncertainty Quantification in Forward and Inverse Problems

Sina Aghaee Dabaghan Fard, Texas A&M University, College Station, TX, United States, Jaesung Lee

Physics-Informed Neural Networks (PINNs) incorporate known physical laws, typically expressed as partial differential equations (PDEs), directly into the training process of deep neural networks. Conventional PINNs produce only point estimates and lack mechanisms for uncertainty quantification, which limits their applicability in settings where knowing model confidence is crucial, such as safety-critical or risk-sensitive situations. Methods such as Bayesian PINNs (B-PINNs) try to address this limitation by applying Bayesian inference across all network layers to model uncertainty. However, a fully Bayesian treatment is often computationally expensive, analytically intractable, and challenging to scale. We propose a Bayesian Last-Layer PINN (BLL-PINN) for solving forward and inverse problems governed by PDEs under uncertainty. By applying a Bayesian prior only to the last layer, BLL-PINN yields an analytically tractable posterior predictive distribution, ensuring uncertainty quantification for both aleatoric and epistemic noise sources. BLL-PINN is a faster yet effective alternative to B-PINN, maintaining scalability and uncertainty quantification without compromising predictive performance. We evaluate BLL-PINN on a range of benchmark problems and compare its performance with state-of-the-art models. Our experiments demonstrate that BLL-PINN achieves competitive predictive accuracy while delivering well-calibrated uncertainty estimates.

3 - Learning Across Machines: Bayesian Meta-Learning for Transferable Process Optimization in Manufacturing

Aditya Satish Rane, Oklahoma State University, Stillwater, OK, United States, Srikanthan Ramesh, Akash Deep

Manufacturing process optimization presents a significant challenge in modern industrial settings, where each machine shows unique operational characteristics despite being manufactured to the same specifications. Traditional optimization approaches optimize each machine independently, resulting in inefficient historical data utilization and prolonged convergence times.

We propose a Bayesian meta-learning framework that enables knowledge transfer across manufacturing machines to accelerate process optimization. Our approach learns shared inductive biases from datasets collected across multiple machines while adapting to new machines with minimal initial experimentation. By leveraging PAC-Bayesian theory, we develop a principled approach for knowledge transfer across machines. The framework employs Gaussian processes as base learners to model complex, non-linear relationships between process parameters and quality metrics, while maintaining crucial uncertainty quantification through Bayesian optimization.

We validate our approach on both synthetic examples and a real-world case study involving FDM-based additive manufacturing and finite element analysis, demonstrating rapid optimization of new setup by utilizing prior experimental data, thus reducing the need to start from scratch each time.

4 - High-Resolution Elasticity Map Estimation from Noisy, Low-Resolution Measurements: An Integrated Inverse Physics-Informed Machine Learning Approach

Tatthapong Srikitrungruang, Texas A&M university, College station, TX, United States, Jaesung Lee

Estimating spatially heterogeneous elasticity maps from low-resolution measurements remains a significant challenge in inverse elasticity problems. Additionally, the presence of substantial measurement noise further exacerbates the complexity. Existing inverse estimation methods typically require high-fidelity data and are highly sensitive to noise. This work presents an integrated Physics-Informed Machine Learning (PIML) framework that fuses spline-based data-driven methods with governing physical laws to robustly reconstruct heterogeneous elasticity distributions, including Young's modulus and Poisson's ratio. The integration of spline methods enhances spatial representation within the PIML framework, enabling effective handling of low-fidelity data. To address measurement uncertainty, we employ a statistical approach based on maximum likelihood estimation to quantify the impact of measurement noise. This advancement holds significant potential for clinical imaging diagnostics and in-situ mechanical characterization. Validation is conducted using simulated two-dimensional displacement data from mechanical testing. The results demonstrate strong robustness to noise and accurate predictions across various data resolutions.

SB63

Building B Level 4 B403

Applications of Forecasting and Reliability

Contributed Session

Contributed

Chair: Viet Nguyen Ca Hoang, University of South Florida, Tampa, FL, United States

1 - Development of a Robust Tool Wear Prediction Method Under Novel Operating Conditions Using Deep Unsupervised Domain Adaptation and Physics-Guided Adjustment

Gyeongho Kim, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea, Republic of, Sang Min Yang, Sujin Jeon, Soyeon Park, Jae Gyeong Choi, Hyung Wook Park, Sunghoon Lim

Data-driven tool wear prediction has become essential in manufacturing as it improves quality and productivity via precise real-time prognosis on machining tool conditions that enable efficient tool utilization and prevent unexpected machining downtime. However, existing tool condition monitoring (TCM) approaches face critical limitations in real-world practice due to frequent changes in operating conditions that cause data distribution shifts (i.e., domain shifts), which make trained models obsolete under novel operating conditions. In addition, since ground-truth tool wear measurements require considerable costs and domain expertise, obtaining labeled data under novel operating conditions is impractical. Given the abovementioned difficulties, this work proposes a deep unsupervised domain adaptation (DA) method with three novel techniques to perform robust tool wear prediction under novel operating conditions. First, inverse Gramian subspace matching (IGSM) is proposed to reduce distribution discrepancy between two different domains by aligning subspaces of inverse Gramians without using labeled samples from novel operating conditions. Second, a pseudo label-based pairwise regularization (PLPR) is developed to transfer informative knowledge regarding tool wear progression to novel operating conditions. Third, a physics-guided adjustment (PGA) is applied during inference to calibrate predictions on unseen tool wear ranges unobserved during training to provide physics-compliant results. Comprehensive experimental results using a real-world milling dataset under various domain shift scenarios demonstrate the proposed method's efficacy in tool wear prediction under novel operating conditions. The proposed method consistently outperforms existing baseline, state-of-the-art DA regression, and state-of-the-art DA-based TCM approaches, indicating its practical effectiveness.

2 - A Feature Fusion Method Using Adjacent-Parameter Ratios for Latent Defect Chip Detection in Semiconductor Fabrication

Young Mok Bae, POSTECH, Pohang, Korea, Republic of, Kwang-Jae Kim

Latent defect chips in semiconductor fabrication present critical product quality and manufacturing efficiency challenges. These latent defect chips exhibit electrical characteristics similar to good chips during the wafer test but fail early during subsequent back-end processes or customer usage. Traditional detection approaches relied on adjusting test voltages based on engineers' subjective judgments. In contrast, recent approaches have shifted toward data-centric methods utilizing electrical parameters collected from the wafer test. As wafer tests have become increasingly complex and detailed, wafer test data volumes have expanded. At the same time, relationships between parameters have gained critical importance for detection performance. In other words, adjacent test parameters contain complementary information that becomes more informative when analyzed as ratios rather than absolute parameter values. Basic classifiers applicable to latent defect chip detection struggle to identify these parameter relationships. The proposed feature fusion method addresses these challenges by deriving electrical parameter ratios and integrating them with raw parameters through a bidirectional cross-attention autoencoder. The autoencoder architecture extracts fused features at configurable dimensions enriched by parameter relationships to enhance latent defect chip detection performance. Experiments on actual wafer test data demonstrate G-mean improvements of up to 6.23 percentage points across basic classifiers, even under strict dimensionality constraints. The domain-informed feature fusion enables test engineers to implement accurate latent defect chip detection with minimal disruption, facilitating proactive quality control and enhanced fabrication efficiency.

3 - Enhancing Reduced-Order Modeling in High-Dimensional Engineering Applications Through Hybrid Gradient Boosting Tree

Melika Baghi, Georgia Institute of Technology, Atlanta, GA, United States, Xiao Liu, Kamran Paynabar

High-fidelity simulations such as computational fluid dynamics (CFD) and finite element analysis (FEA) are essential for modeling complex engineering systems but remain prohibitively expensive for many-query tasks including parametric studies, optimization, and real-time control. Projection-based reduced-order models (ROMs), most often constructed via Proper Orthogonal Decomposition (POD), alleviate this cost by projecting full-order dynamics onto low-dimensional subspaces. However, conventional POD-Galerkin ROMs can lose accuracy under parameter variation, struggle in nonlinear or discontinuous regimes, and lack mechanisms to preserve geometric structure or quantify predictive uncertainty.

This work introduces a hybrid framework that integrates Gradient Boosting Trees (GBT) with a geometry-preserving embedding of POD subspaces on the Grassmann manifold. A smooth, injective mapping from the manifold to a compact Euclidean space enables efficient GBT regression while maintaining the intrinsic geometry of the bases, and a norm-constrained ensemble training strategy enforces orthonormality and injectivity in predictions.

Validation on two benchmark problems on unsteady viscous flow past a cylinder and compressible shock tube dynamics demonstrates that the proposed method outperforms interpolation-based ROMs, achieving higher accuracy and robust generalization across nonlinear parameter spaces. These results show that combining geometric learning with ensemble regression offers a scalable, interpretable, and physically consistent pathway for reduced-order modeling in high-dimensional, parameter-varying PDE systems.

4 - Modeling of Shape Populations in Manufacturing: An Expectation-Maximization Approach on Riemannian Manifolds

Rifat Hasan, Florida A&M University, Florida State University (FAMU-FSU) College of Engineering, Tallahassee, FL, United States, Chiwoo Park, Arda Vanli

Geometrical forms of products serve as very important quality characteristics in advanced manufacturing, e.g., geometries of 3D printed products and morphology of nano-manufactured products. Quantifying the distribution of product's geometrical forms is critical for assessing both product quality and process variation associated with manufacturing processes. In this study, we propose a new statistical model to characterize the probability distribution of geometric forms in two dimensions. Based on the Riemannian geometry, we define an appropriate space of shapes and proper metric over the space. On the metric space, we model a Gaussian-type distribution and a mixture of the distributions. We have also introduced the multimodality of shapes in the mixture of the distributions. We also present the EM-type algorithm to infer the distribution with a finite number of shape samples from processes. The numerical results of the algorithm are presented with simulation cases and applications to modeling the shape distribution of nanoproducts.

5 - Grid-Scale Battery Life Prediction Using Equivalent Circuit and Particle-Filtered Stochastic Modeling

Krishna Rijal, University of South Florida, Tampa, FL, United States, Mingyang Li, Tapas Das

Increasing penetration of renewables and decreasing cost of lithium-ion batteries make large-scale battery energy storage systems (BESS) a promising solution for supporting grid balance and reliability, and a resource for arbitrage. This is evident from the increasing number of BESS that are either being installed or planned. However, battery systems suffer from degradation from usage, significantly impacting their performance over time. Proper characterization of battery degradation and accurate prediction of the remaining useful life of BESS components are ongoing issues that are yet to be fully addressed. These issues are of critical interest to the BESS operators, whose primary goal is to ensure effective and reliable operation of the system. We propose a new methodology for online detection of battery degradation and prediction of the remaining life using an equivalent electrical circuit model. The circuit parameters are identified and filtered, and with this information, a random process mimicking battery health progression is simulated. The simulation results provide estimates of the first hitting time and the mean hitting time to reach the battery life threshold.

6 - Sensor Placement for Anomaly Detection in Large Scale Networks Under Limited Budget

Feiran Xu, University of Iowa, Iowa City, IA, United States

As large-scale graph-structured systems become increasingly common, maintaining their stability and reliability is critical. A major challenge lies in promptly detecting anomalous nodes and tracing the origins of disturbances. While real-time sensor data from within the network can aid in this process, deploying sensors at every node is often impractical due to cost and logistical constraints. This makes the strategic placement of a limited number of sensors essential for maximizing system observability. In this work, we propose a novel cluster-based framework for selecting optimal sensor locations to support anomaly detection in large networks. Through a series of data-driven experiments, we demonstrate that our approach effectively enhances network observability while operating within budget limitations.

7 - Mechanism-Aware Validation of Data-Driven Cardiovascular Phenotypes in Obstructive Sleep Apnea Using Structural Equation Modeling

Viet Nguyen Ca Hoang, University of South Florida, Tampa, FL, United States, Khanh Do, Phat Huynh, Abhay Sharma, Trung Le

Obstructive Sleep Apnea (OSA) is associated with elevated cardiovascular disease (CVD) risk, yet existing risk models often overlook underlying phenotypic heterogeneity driven by distinct risk pathways. We propose a hybrid modeling framework that integrates outcome-guided feature selection, unsupervised clustering, and structural equation modeling (SEM) to identify and clinically validate OSA-CVD phenotypes.

Using data from 633 patients with 121 multimodal features spanning polysomnographic, clinical, and biomarker domains, we first apply outcome-specific models targeting major adverse cardiac events (MACEs) and apnea-hypopnea index to extract features predictive of CVD and OSA severity. These features are used in hierarchical clustering, evaluated by 5-fold stratified train/validation splits, internal clustering metrics, and external MACE associations.

To clinically validate these data-driven phenotypes, we construct domain-informed SEMs reflecting hypothesized OSA-CVD pathways (e.g., sleep disruption to inflammation to CVD) and test for multi-group path invariance. Unlike conventional phenotyping approach, our framework enables formal hypothesis testing of biological mechanisms underlying observed subgroups. SEM complexity is constrained by a sample-size-to-parameter ratio $\geq 10:1$.

Preliminary findings suggest at least three distinct phenotype clusters with differential CVD risk profiles. We also report bootstrapped confidence intervals for MACE odds ratios across clusters. This approach enables both quantitative validation (via model fit, path invariance, and outcome prediction) and qualitative assessment grounded in clinical understanding.

This mechanism-aware pipeline facilitates interpretable phenotype discovery and generalizes to other multimorbid conditions with known mechanisms and multimodal data. By uncovering phenotype-specific pathways, it may inform targeted treatments, support earlier risk stratification than outcome-based methods, and improve personalized CVD care in OSA.

SB64

Building B Level 4 B404

Behavioral Operations Management-Best Paper Competition

Award Session

Behavioral Operations Management

Chair: Chenshan Hu, Washington University in St. Louis, Saint Louis, MO, United States

1 - How a Study of Wait Utility Became a Study of Presentation Format: People Chose Queues over Clocks, 80&Ndash;20

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

Customer utility depends on wait characteristics, which are shaped by the decisions of system designers. To inform these decisions, we examine how four fundamental wait attributes—average duration, variability, reward for completion, and line length—affect utility and reveal the overlooked role of wait presentation. Specifically, we measure how customer behavior responds to changes in wait format (traditional Queues vs. countdown Clocks) and in the amount of information disclosed (average duration, variability, or no information at all). Study 1 employs conjoint analysis, incentivized with a real wait, to estimate the relative magnitudes of utility across fundamental wait attributes. People dislike long and variable waits and value large rewards for completion. Between presentation schemes, participants behave as though their waiting cost is approximately twice as high when waits are presented as Clocks rather than Queues. Study 2 asks participants to estimate the duration of given waits; they underestimate mean durations by 20% when presented as Queues but estimate them accurately when presented as Clocks. Study 3 asks participants to choose which wait they will endure in real time between two equivalent waits that differ only in presentation format. Nearly 80% prefer the Queue to the Clock, consistent with the model-based prediction from Study 1. This research highlights an underappreciated managerial lever: presentation format. It can reduce disutility without requiring operational redesign.

2 - Pricing and Addiction Control for Digital Services

Xiao Lei, University of Hong Kong, Hong Kong, Hong Kong

The rising use of digital services like video games and video platforms has led to concerns about digital addiction. Unlike traditional addictive goods, these services often use personalized, engagement-based pricing and offer highly varied user experiences. While empirical and behavioral research on digital addiction is growing, a theoretical understanding of how pricing and policies influence addiction is lacking. This study provides the first analytical model of pricing strategies and control measures for addictive digital services, aiming to guide firms and policymakers in addressing addiction without sacrificing profitability. Using dynamic programming to model habit-forming user behavior, we characterize and compare optimal fixed and dynamic pricing. Our findings show that dynamic pricing not only increases profits but also reduces long-term addiction, creating a win-win scenario. One-time new-user promotions, while initially increasing addiction, lead to a net reduction over time. We also evaluate three addiction control policies: consumption limits, price floors, and consumption taxes. The analysis of their profit-addiction efficiency reveals that a consumption limit or tax is most effective, with the variation in user experience being a key determining factor. A numerical case study demonstrates that, compared to fixed pricing, dynamic pricing alone can increase profit by 3.07-3.73% while reducing addiction by 1.31%. When combined with additional controls, it can achieve a 20.65-22.16% reduction in addiction with no profit loss.

3 - Biases of Humans, of AI, and of Humans with Ai

Anton Ovchinnikov, Smith School of Business, Queen's University, Kingston, ON, Canada

Prior research examined whether modern AIs—such as large language models (LLMs) like ChatGPT—are themselves biased. This paper asks a deeper question: how do human decision biases evolve when people interact with biased or unbiased LLMs? We introduce a novel interactive AI experimental paradigm in which human participants complete decision tasks under four conditions: alone, with a human-like AI, an unbiased AI, or with an AI deliberately calibrated to differ in bias. We test this framework across three canonical behavioral operations biases. The results show: (1) AI bias measurably shifts human decision patterns, amplifying or attenuating human bias depending on the AI's stance; (2) a previously undocumented "agentic nudging" effect, in which controlled system prompts steer human choices in predictable directions; and (3) a refined account of algorithm aversion: that the human-AI interaction—not generic distrust of AI—determines collaboration outcomes, either improving or degrading performance.

4 - Video-Based AI Misconduct Monitoring: Unintended Consequences on Operational Outcomes

Zhanzhi Zheng, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Employee misconduct presents ongoing challenges for organizations, yet traditional monitoring methods, such as CCTV, are largely reactive and offer little chance for timely intervention. Recent advances in computer vision and edge computing have enabled AI-enhanced video monitoring systems capable of detecting misconduct in real time and providing immediate feedback—an innovation not previously feasible. Despite its growing adoption, the operational implications of this technology remain underexplored.

Using a unique dataset from a restaurant chain that implemented video-based AI misconduct monitoring, we examine the effects of real-time feedback on organizational performance. Contrary to expectations, we find that deploying this technology significantly reduces performance, as measured by revenue and customer ratings. Further analysis reveals that this effect is primarily driven by high-quality restaurants, while lower-quality restaurants may experience potential improvement.

We also show that the type of feedback matters: individual-focused feedback (e.g., uniform violations) intensifies negative effects, whereas process-focused feedback (e.g., unattended cashier stations) mitigates performance declines.

Our findings highlight the unintended consequences of adopting advanced AI surveillance systems and emphasize the importance of organizational context and feedback design. By illustrating how real-time monitoring can produce counterintuitive outcomes, this study offers new managerial insights into the complex relationship between AI adoption, employee behavior, and operational performance.

SB65

Building B Level 4 B405

Decision-Making with Generative AI

Invited Session

Artificial Intelligence

Chair: Yan Leng, University of Texas at Austin

Co-Chair: Leqi Liu, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Leveraging Large Language Models for Collective Decision-Making

Marios Papachristou, Cornell University, Ithaca, NY, United States, Longqi Yang, Chin-Chia Hsu

In various work contexts, such as meeting scheduling, collaborating, and project planning, collective decision-making is essential but often challenging due to diverse individual preferences, varying work focuses, and power dynamics among members. To address this, we propose a system leveraging Large Language Models (LLMs) to facilitate group decision-making by managing conversations and balancing preferences among individuals. Our system aims to extract individual preferences from each member's conversation with the system and suggest options that satisfy the preferences of the members. We specifically apply this system to corporate meeting scheduling. We create synthetic employee profiles and simulate conversations at scale, leveraging LLMs to evaluate the system performance as a novel approach to conducting a user study. Our results indicate efficient coordination with reduced interactions between the members and the LLM-based system. The system refines and improves its proposed options over time, ensuring that many of the members' individual preferences are satisfied in an equitable way. Finally, we conduct a survey study involving human participants to assess our system's ability to aggregate preferences and reasoning about them. Our findings show that the system exhibits strong performance in both dimensions.

2 - Dbot: Artificial Intelligence for Systematic Long-Term Investing

João Sedoc, New York University, New York, NY, United States, Vasant Dhar

Long-term investing was previously seen as requiring human judgment. With the advent of generative artificial intelligence (AI) systems, automated systematic long-term investing is now feasible. In this paper, we present DBOT, a system whose goal is to reason about valuation like Aswath Damodaran, who is a unique expert in the investment arena in terms of having published thousands of valuations on companies in addition to his numerous writings on the topic, which provide ready training data for an AI system. DBOT can value any publicly traded company. DBOT can also be back-tested, making its behavior and performance amenable to scientific inquiry. We compare DBOT to its analytic parent, Damodaran, and highlight the research challenges involved in raising its current capability to that of Damodaran's. Finally, we examine the implications of DBOT-like AI agents for the financial industry, especially how they will impact the role of human analysts in valuation.

3 - The Habermas Machine: AI-Mediated Systems for Collective Human Decision Making

Michiel Bakker, MIT, Boston, MA, United States

As AI systems become more powerful, there is an urgent need for new systems and institutions that support collective human oversight and decision making. We present a set of AI-mediated systems that combine human input with large language models to help communities deliberate, resolve disagreements, and make accurate collective judgments. These systems, presented in NeurIPS and Science scale democratic processes through model-assisted deliberation, collective fact-checking and oversight of AI systems. First, we introduce the "Habermas Machine", which refines group statements based on participants' opinions and critiques. In a preregistered study (N = 5,734), participants preferred AI-mediated statements over those written by human mediators and often updated their own views. These results were replicated in a demographically representative citizens' assembly. Second, we extend the approach to Community Notes-style fact-checking systems, using LLMs and reward models to synthesize diverse viewpoints into helpful explanatory notes. Finally, we propose methods to align incentives: LLM aggregators use crowd-submitted rationales to form

predictions and distribute rewards based on marginal value. Together, these systems demonstrate how LLMs can support scalable, pluralistic oversight in high-stakes domains.

4 - Active Exploration via Autoregressive Generation of Missing Data

Tzu-Ching Yen, Columbia University, New York, NY, United States

We pose uncertainty quantification and exploration in online decision-making as a problem of training and generation from an autoregressive sequence model, an area experiencing rapid innovation. Our approach rests on viewing uncertainty as arising from missing future outcomes that would be revealed through appropriate action choices, rather than from unobservable latent parameters of the environment. This reformulation aligns naturally with modern machine learning capabilities: we can i) train generative models through next-outcome prediction rather than fit explicit priors, ii) assess uncertainty through autoregressive generation rather than parameter sampling, and iii) adapt to new information through in-context learning rather than explicit posterior updating. To showcase these ideas, we formulate a challenging meta-bandit problem where effective performance requires leveraging unstructured prior information (like text features) while exploring judiciously to resolve key remaining uncertainties. We validate our approach through both theory and experiments. Our theory establishes a reduction, showing success at offline next-outcome prediction translates to reliable online uncertainty quantification and decision-making, even with strategically collected data. Semi-synthetic experiments show our insights bear out in a news-article recommendation task, where article text can be leveraged to minimize exploration.

5 - Tell Me Why: Incentivizing Explanations

Siddarth Srinivasan, Harvard University, Cambridge, MA, United States, Ezra Karger, Michiel Bakker, Yiling Chen

Effective human-AI collaboration requires mechanisms that incentivize truthful, informative contributions, especially when interactions occur through natural language. We introduce a novel collective intelligence mechanism inspired by prediction markets. Unlike traditional prediction markets which primarily elicit numerical forecasts, our mechanism elicits human-generated natural-language *explanations* for beliefs. We formally show that this mechanism is incentive-compatible, meaning providing honest and informative explanations constitutes a Bayesian Nash equilibrium under our design.

A simple instantiation of this mechanism involves participants submitting natural-language rationales responding to a question posed by a mechanism designer (e.g., predictions about policy outcomes or future events). A large language model (LLM) supervisor assesses each rationale, updates its internal beliefs accordingly, and maintains a collaboratively updated wiki-style resource. Contributors receive rewards proportional to how significantly their rationales alter the LLM's beliefs towards some eventually-revealed reference answer.

We focus on applying our mechanism to judgmental forecasting tasks in particular, where experts make forecasts about uncertain future events to aid decision-making. Forecasting provides a clear decision-making application as the outcome of future events offer objective ground truth resolution for evaluating rationales.

SB66

Building B Level 4 B406

Optimization Models for University Operations

Invited Session

INFORMS Committee on Teaching and Learning

Chair: Richard Forrester, Dickinson College, Carlisle, PA, United States

Co-Chair: Lucas Waddell, Bucknell University, Lewisburg, PA, United States

1 - Assigning Faculty in University Committee Service

Liz Bouzarth, Furman University, Greenville, SC, United States, Chelsea Alford, Ben Grannan, Kevin Hutson, Ella Morton

Institutional expectations regarding faculty service vary, but at many schools, faculty members are expected to serve in various roles outside of teaching and research. This often leads to real or perceived differences in service load through both official and unofficial duties like advising, recruiting, departmental service, university committee service, and service to the profession beyond the institution. This talk provides a case study of an integer programming model built to help assign faculty to standing committee service at Furman University based on optimizing faculty preferences while meeting constitutional committee requirements. The annual task of assigning faculty to openings on university committees had previously been done manually, but the use of the integer programming formulation has helped seed the committee's work in creating the annual committee slate for the past five years. Observations regarding service equity will be shared and generalized.

2 - Building a Framework to Schedule Courses at Community Colleges Based on Student Success by Using Binary Integer Programming

Fabio Vitor, University of Nebraska at Omaha, Omaha, NE, United States, Joseph Bonacci

Community colleges play a vital role in the U.S. higher education system by offering both technical training and pathways to bachelor's degrees. These institutions are crucial in shaping the current and future workforce. One notable feature of community colleges is their flexible scheduling, which is designed to accommodate students' varying needs. Effective scheduling practices are essential to ensure accessibility and success for all students. To address this, it is important to consider several factors when scheduling courses, though balancing these factors can be challenging, particularly for departments with limited resources. This talk will discuss an optimization framework for scheduling courses at community colleges with a focus on enhancing student success. The model employs binary integer programming and addresses key elements that contribute to student success, including course frequency and timing, delivery modes (in-person, online, and hybrid), instructor preferences and workloads, integration of high school students, room and technology appropriateness, and co-taught courses. The model is applied and evaluated using data from the Mathematics Department at a community college in Iowa. This optimization framework has the potential to assist community colleges in creating schedules that not only promote student success but also further their goals of accessibility and equity.

3 - Optimizing Final Exam Schedules at Bucknell University

Lucas Waddell, Bucknell University, Lewisburg, PA, United States, Samuel Gutekunst, Thiago Serra, Clara Chaplin, Stanley Gai, Tsugunobu Miyake, Luke Snyder, Vy Tran

This talk will introduce an interdisciplinary undergraduate research project concerning the creation of optimized final exam schedules at Bucknell University. This project, which involves data visualization, integer programming, and web-app development, resulted in an open-source tool that has been used by the Bucknell Registrar's office to schedule final exams since the Spring 2024 semester.

4 - The First-Year Seminar Assignment Problem: a Multi-Objective Optimization Approach

Richard Forrester, Dickinson College, Carlisle, PA, United States

Like many colleges, Dickinson College uses first-year seminars to help students transition into college-level work through writing, research, and discussion. Students rank their top seminar choices from a diverse list, and the college assigns them to seminars with the dual objectives of honoring individual preferences and maintaining diversity within each seminar. While students prefer placement in one of their higher-ranked seminars, faculty aim to ensure that seminar demographics are roughly representative of the entire student body, creating a microcosm of the college. In this talk, we describe a student-faculty research project focused on navigating the trade-offs inherent in seminar assignments by modeling the problem as a multi-objective optimization task. This work demonstrates how mathematical modeling can support institutional decision-making while providing meaningful research experiences for undergraduate students.

SB67

Building B Level 4 B407

Data Mining Best Paper Award

Award Session

Data Mining

Chair: Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

1 - Supervised Clustered Interpretability: a Novel Method Towards Transparent AI with Social Good Implications

Matt Baucum, Oklahoma State University, Stillwater, OK, United States, Meysam Rabiee

This paper introduces clustered interpretability, a novel approach enhancing AI transparency and subgroup fairness. Our innovative Separation-Disparity (SD) metric, optimized via tailored particle swarm optimization (PSO), ensures robust subgroup homogeneity. Comprehensive validation using synthetic and real Parkinson's disease datasets demonstrates methodological soundness, practical effectiveness, and novel contributions to interpretable AI literature.

2 - Joint and Condition Modeling for High-Dimensional Aleatoric and Epistemic Uncertainty Quantification

Yongxiang Li, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Chunlin Ji, Dingwei Gong, Meiying Zhang, Terry Ma, Jun Liu

Uncertainty quantification is vital for modeling and decision-making. We propose a scalable approach that jointly models input-output distributions via Gaussian mixtures, enabling closed-form posterior inference. A Stiefel-constrained projection identifies low-dimensional features, optimized via Riemannian SGD. Our method captures aleatoric and epistemic uncertainties efficiently in high-dimensional, data-limited settings.

3 - First-Order Policy Optimization for Robust Policy Evaluation

Yan Li, Texas A&M University, College Station, TX, United States, Guanghui Lan

(Distributionally) robust reinforcement learning aims to learn policies resilient to environment shifts. Yet it is well known that existing methods using function approximation suffer from provable divergence. We resolve this open problem by developing the first robust policy evaluation framework that supports function approximation with provable global convergence, while attaining optimal iteration and sample complexities.

4 - Distribution-in-Distribution-Out Regression

Xiaoyu Chen, University at Buffalo, Buffalo, NY, United States, Yujing Huang, Xinwei Deng

A distribution-in-distribution-out (DIDO) regression is proposed and validated by defining arithmetic operations among probability distributions based on the differential geometry of pseudo Riemannian manifold. We show that the Fréchet least squares is the best linear unbiased estimator, building the foundation for estimation and inference when random objects are probability distributions.

SB68

Building B Level 4 B408

Frontiers of Machine Learning for Autonomous and Intelligent Manufacturing Systems

Invited Session

Data Mining

Chair: Intiaz Ahmed, West Virginia University, Morgantown, United States

Co-Chair: Yinsong Wang, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Yinsong Wang, Georgia Institute of Technology, Atlanta, GA, United States

1 - Automatic Melt Pool Tracking and Segmentation in Laser Powder Bed Fusion Using X-Ray Image Sequence

Ruiyuan Zhang, Virginia Tech, Blacksburg, VA, United States, Zhenyu Kong, Maede Maftouni, Jihoon Chung, Bo Shen, Rongxuan Wang, Andrew Law, Prahalada Rao

Laser powder bed fusion (LPBF) is one of the most extensively used techniques for metal additive manufacturing (AM). This process employs a laser to melt metal powder onto a substrate, creating a melt pool. The solid-liquid interface of the melt pool plays a critical role in determining the cooling process, which, in turn, influences the microstructure and the mechanical properties of the printed part. High-speed X-ray imaging has been utilized to study subsurface melt pool phenomena in real-time. However, segmenting the melt pool from X-ray images is challenging due to high noise levels and low contrast. Despite the widespread application of high-speed X-ray imaging, efficient data processing techniques remain underdeveloped. Researchers often rely on manual annotation or basic image processing techniques for object segmentation. These methods are either time-consuming or lack the necessary accuracy and robustness. This work presents a video object segmentation deep learning model to automatically track and segment the melt pool in X-ray image sequences. The proposed model is semi-supervised, requiring only the manual annotation of the melt pool boundary in the first frame to predict it for the subsequent frames. It incorporates spatio-temporal attention modules to effectively capture correlations in X-ray image sequences. Specifically, the co-attention block captures temporal features from the previous frame, while the attention block localizes the key regions within the current frame. Experimental results demonstrate that incorporating attention modules improves melt pool segmentation accuracy compared to existing state-of-the-art methods.

2 - Enhancing Surprise-Based Experimentation with Pointwise Mutual Information for Intelligent Manufacturing

Quan Zeng, Georgia Institute of Technology, Atlanta, GA, United States, Yu Ding, Yinsong Wang

We propose a decision-making framework that integrates pointwise mutual information (PMI) to guide autonomous experimentation. We incorporate PMI into both exploration and exploitation phases to enable informed and locally adaptive search behavior. PMI provides a localized view of the input-response landscape, helping the system identify regions with high expected information gain. We will discuss what differences the PMI guided policy can make.

3 - Privacy-Preserving Porosity Prediction Using Hierarchical Graph-Attention Networks in Laser-Based Additive Manufacturing

Mahathir Mohammad Bappy, Louisiana State University, Baton Rouge, LA, United States

Ensuring structural integrity in metal additive manufacturing (MAM) is critical for aerospace, biomedical, and defense applications where even microscopic defects (i.e., porosity) can lead to catastrophic failures. Porosity control in the LENS process is particularly vital, as internal pores compromise the mechanical strength of printed parts. However, traditional predictive models often struggle to capture the complex, layer-by-layer connectivity and nuanced relationships embedded in frame-level metadata, resulting in suboptimal porosity predictions. To address this challenge, we propose a Hierarchical Graph Attention Network (HGAT) that models intra-layer dependencies while preserving global context. The approach leverages a unified graph based representation of multimodal data, including process parameters and high resolution thermal imagery. Each melt-pool observation is treated as a node in the graph, with edges encoding spatial proximity through k-nearest neighbor connections. This structure enables the HGAT to extract intricate patterns that govern porosity formation. Recognizing the growing need for secure data sharing among distributed suppliers, we further integrate a Feature-Importance Local Differential Privacy mechanism. By applying importance-weighted Gaussian noise, our framework achieves (ϵ, δ) -LDP guarantees while maintaining critical feature fidelity. Evaluation on a public Ti-6Al 4V thin-wall dataset shows that the private HGAT achieves an F1 score of approximately 0.94 at $\epsilon = 1.0$, a performance drop of less than 2% compared to the non-private baseline ($F1 \approx 0.96$) while sustaining a balanced accuracy above 0.92 across various privacy settings. These findings highlight the feasibility of accurate, privacy-preserving porosity prediction, offering a viable path toward secure data exchange and quality assurance.

4 - End-2-End Resilience for Machine Tool Controller Against Cyberattacks

Shashank Galla, Texas A&M University, College Station, TX, United States, Abhishek Hanchate, Vaishnavi Garudkar, Yuhao Zhong, Satish Bukkapatnam

Cyberattacks on sensors and actuators are surging in manufacturing cyber-physical systems (CPS), mainly on machine tool controllers (MTCs). These attacks compromise product quality and disruption schedules. Manufacturing systems must ensure end-to-end resilience to address these threats, which requires fast detection, stable recovery,

and effective mitigation. We propose an integrated framework that combines Neural Network (NN)-based attack detection with Reinforcement Learning (RL)-based control for recovery and mitigation of MTCs under cyberattack scenarios. The framework's effectiveness is validated through two stepper-motor case studies: one in a digital-twin simulation and one with physical hardware. Our results confirm the framework's ability to detect attacks swiftly, maintain system stability, and minimize overall damage.

5 - Domain-Aware Denoising Diffusion Probabilistic Model with Class Conditioning for Melt Pool Minority Class Augmentation

Intiaz Ahmed, West Virginia University, Morgantown, WV, United States, Israt Zarin Era, Srinjoy Das, Zhichao Liu

Thermal imaging of melt pools in laser-based additive manufacturing presents a critical challenge in anomaly detection due to severe class imbalance, where normal instances vastly outnumber abnormal ones. This imbalance impairs the generalization ability of supervised classifiers. To address this, we propose a data augmentation framework using a Denoising Diffusion Probabilistic Model (DDPM) to synthesize high-fidelity thermal images for the minority class. The DDPM captures complex spatial and intensity patterns of melt pool dynamics through iterative noise-to-image denoising. We introduce two key innovations: (1) domain-aware spatial priors based on thermal intensity weighting to emphasize melt pool regions during training, and (2) class conditioning via a context encoder, which injects label-specific information into the bottleneck of the DDPM's U-Net backbone. The context encoder transforms discrete class labels into spatially aligned feature maps that guide the generative process by fusing with bottleneck features. This enables the model to generate class-specific thermal patterns while preserving local structure and physical fidelity.

SB69

Building B Level 4 B409

Managing in Innovative Environments

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: John Angelis, University of Lynchburg, Lynchburg, VA, 24504, United States

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

1 - Organizational Learning and Ambidexterity in a Multi-Level, Multi-Generation, Multi-Phase Process Development Effort

Charles Weber, Portland State University, Portland, OR, United States, Timothy Anderson

A qualitative study of process development in the semiconductor manufacturing ecosystem reveals that organizational learning in this environment is multidimensional and presents the developers with severe ambidexterity challenges. Small teams of developers must concurrently develop constituent technologies for three process generations at three different stages of maturity and multiple levels of integration on the same pieces of equipment. Thus, the need to exploit comes in conflict with the need to explore. Overcoming the ambidexterity challenge involves making knowledge more explicit, as well as designing an organization whose structure exhibits fidelity with the knowledge it attempts to acquire.

2 - Using Genai as An Expert: a Study of Power Generation in Remote Locations

Mark Ryan, Portland State University, Beaverton, OR, United States, Charles Weber

A study of power generation in remote locations, which is based on the opinions of experts, finds that these experts may be hard to find. The study explores how GenAI can be used as an expert when human experts are scarce.

3 - The Effects of Personal Traits on Revisit Likelihood in Data Breaches

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

Data breaches are increasingly common events that undermine customer trust, disrupt consumer-firm relationships, and generate long-term reputational risks. Prior research has examined how consumers assign blame for breaches, how breach severity influences revisit intentions, and how emotional reactions such as fear or anger shape post-breach behaviors. Drawing on attribution theory, protection motivation theory, consumer ethnocentrism theory, justice theory, and risk perception theory, this study extends the "Better Angry than Afraid" emotional processing framework to investigate how personal traits and breach characteristics interact to influence post-breach decision-making.

Specifically, we explore the roles of ethnocentrism, political ideology (conservative vs. liberal), and the experience of multiple breaches from the same company in shaping revisit intentions. Using a scenario-based survey with a diverse sample of U.S. consumers, we manipulate breach recurrence and measure trait-level variables alongside demographic, educational, and behavioral factors, including cybersecurity knowledge and personality characteristics. By integrating perspectives from consumer behavior, political psychology, and information security, this research seeks to advance understanding of how individual predispositions, perceived fairness, and risk assessments jointly shape consumer responses, contributing to a more comprehensive framework for trust recovery and customer relationship management in the aftermath of data breaches.

4 - Tackling Decomposition and Dynamics of Innovation

Juliana Hsuan, Copenhagen Business School, Frederiksberg, Denmark

Decomposition is a process that can facilitate the breaking down of complex systems so standardization and innovation can be managed independently. As systems advance and grow for economies of scale and scope, the tradeoffs and nuances that arise from such process are not trivial. This paper explores how tradeoffs can be operationalized and measured.

5 - Leveraging Weighted Syntactic and Semantic Attention Summary (Ssas) Towards Text Categorization Using LLMs

Nitin Joglekar, Villanova School of Business, Villanova, PA, United States, Nitin Mayande, Sharookh Daruwalla, Shreeya Verma Kathuria, Charles Weber

Automatic summarization, a subfield of NLP, tackles the challenge of efficiently extracting key points from vast amounts of text. However, achieving accurate summaries requires models to go beyond simple word frequencies. Accurate summarization hinges on effectively bridging the gap between syntax and semantics. Our proposed methodology, Syntactic and Semantic Attention Summary (SSAS), uses a combined approach to provide a more nuanced understanding of the text by generating summaries that are not just grammatically sound but also semantically accurate. While SSAS does an excellent job in capturing the relevant information within large text data, the methodology can be improved. Weighting the data points is one such improvement - it provides a way to give a higher weight to information that is seen more frequently across the dataset while simultaneously pruning out less valuable information. This enables the summary generated from Weighted-SSAS (wSSAS) to be more representative of the main information contained in a dataset. One useful application of wSSAS-based summarization is data categorization. Data categorization enables bucketing individual datapoints into either pre-defined or auto-generated categories with the help of LLMs. In this study, we discuss the impact of weights on data summarization and ultimately, on the task of data categorization.

Sunday, October 26, 1:15 PM - 2:30 PM

MIF Poster

Building A Level 1 Exhibit Hall A3

MIF Poster Competition

Award Session

Minority Issues Forum

Chair: Karen Hicklin, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, United States

1 - Nonlinear Spatio-Temporal Run-to-Run Control: Leveraging Diffusion Models for Disturbance Management**Zihan Zhang, Georgia Institute of Technology, Atlanta, GA, United States**

Run-to-run (R2R) control is widely used in advanced manufacturing processes. Conventional R2R control methods often assume that system disturbances exhibit linear structures or low-rank approximations in high-dimensional data, such as images or videos. While these assumptions facilitate implementation, they limit the effectiveness of R2R control in managing sophisticated manufacturing processes with intricate nonlinear disturbances. To address these challenges, we propose a novel nonlinear spatio-temporal R2R control framework that estimates system disturbances using a diffusion model and applies control actions to compensate for them. The proposed approach integrates an offline modeling phase to capture disturbance dynamics and an online control phase that dynamically adjusts actions to minimize deviations from the target system response, ensuring system stability and precision. Unlike conventional deterministic control strategies, the proposed method explicitly quantifies uncertainty, enabling more robust and informed decision-making. The effectiveness of this method is validated through simulation studies and a case study, demonstrating its adaptability in complex, high-dimensional environments.

2 - An Equitable Energy Transition: Integrating Community Perspectives into Power System Modeling in An Emerging Economy - Poster**Faustina Sarfo, UMass Amherst, Amherst, MA, United States**

This project integrates community-based participatory research into energy modeling to address equity gaps in Ghana's power system. We co-develop equity-focused metrics with underserved communities and implement them in GenX, a capacity expansion model. By linking system cost, reliability, and access, we evaluate energy pathways that reflect local priorities and support inclusive, just energy transitions.

3 - Dynamic Transmission Line Switching Amidst Wildfire-Prone Weather Under Decision-Dependent Uncertainty**Juan-Alberto Estrada-Garcia, University of Michigan, Ann Arbor, MI, United States**

Dry and windy seasons increase wildfire risk conditions, threatening transmission lines and causing grid vulnerabilities. We propose a multi-stage optimization model that adaptively adjusts grid topology to optimally respond to wildfire propagation, accounting for decision-dependent uncertainty in which line survival probabilities depend on usage. Using distributionally robust optimization and a stochastic nested decomposition variant, we provide a finite convergence guarantee and show strong performance on realistic California grid data against two-stage policies.

4 - A Novel Method to Identify Potential Collaborators with A Patent-Behavior Knowledge Graph**Tongrui Zhang, Sun Yat-sen University, Guangzhou, China, People's Republic of**

This study proposes a framework for constructing an innovation collaboration knowledge graph by integrating multilayer patent collaboration behaviors. It transforms the static network into a dynamic and predictive one, enhancing path strength by 428.06% and collaborator recommendations by 57.14%. The framework enriches the innovation collaboration network theory by offering a dynamic perspective and supports data-driven collaborator identification to reducing bias in managerial decision-making.

5 - Finding Minimal Monitoring Windows for Fetal Risk Prediction: a Two-Stage Machine Learning Framework**Jaber Qezelbash-Chamak, University of Florida, Gainesville, FL, United States**

Prolonged fetal heart rate (FHR) monitoring complicates timely and accurate anomaly detection due to variability. We propose a two-stage machine learning approach to optimize the minimal CTG monitoring duration required for reliable fetal anomaly detection, achieving accuracy, precision, recall, and specificity exceeding 99% and enhancing clinical decision-making efficiency.

6 - Leveraging Inequality Measures for Equitable Food Bank Operations: a Data-Driven Approach**Enoch Bonsu, North Carolina A&T State University, Greensboro, NC, United States**

This study employs the Gini and Hoover indices to assess equity in food distribution at a local food bank. Using Meals Per Person in Need data, we reveal rising network inequality and localized branch imbalances. These standardized measures provide a quantitative framework for identifying disparities, tracking trends, and guiding resource allocation, thereby advancing equity in food banking.

7 - Modeling Spatial Health Damage Disparities from CCS Under Atmospheric and Policy Uncertainty - Poster**Nexus Attiogbe, UMass Amherst, Amherst, MA, United States**

Recent U.S. policies, including the Inflation Reduction Act's 45Q Tax Credit and the EPA's CO₂ Performance Standards, aim to incentivize carbon capture (CC) in the power sector. This study evaluates how these individual policies and their combinations affect CC deployment, co-pollutant emissions (NO_x, SO_x, PM_{2.5}, NH₃, and VOCs), and the distribution of health damages across socioeconomic groups. Using a stylized three-node model, we optimize the generation investment and dispatch. We apply this regional case studies in Alabama and Texas to investigate health disparities. Air pollution exposure modeling is paired with census tract-level data to assess disparities in per-capita monetized health damages. Regional variation in co-pollutant dispersion patterns due to atmospheric conditions is incorporated to reflect climate-based uncertainties. This study highlights the importance of considering environmental injustice when designing policy by analyzing trade-offs between CC adoption, co-pollutant exposure, and equity outcomes. Results emphasize the need for decarbonization strategies that remain robust under uncertainty and prioritize equitable emission reductions.

8 - Differentially Private Sequential Learning - Poster**Yuxin Liu, University of Pittsburgh, Pittsburgh, PA, United States**

In a differentially private sequential learning setting, agents introduce endogenous noise into their actions to maintain privacy. We develop a smoothed randomized response mechanism for continuous signals, achieving faster learning ($\Theta(\epsilon \log(n))$) and finite stopping time—compared to the slower $\Theta(\sqrt{\log(n)})$ rate in non-private regimes. For binary signals, privacy can help avoid early information cascades and increase the probability of a correct cascade.

9 - Bridging the Gap: Quantifying Spatial and Temporal Access Within a Food Bank Network**Rashik Siddiquee, NCSU, Raleigh, NC, United States**

This study presents a data-driven framework to quantify accessibility to food pantries, focusing on food-insecure populations and varying transportation access. By integrating geospatial data on food pantries, transit stops, road networks, and food insecurity rates, we derive tract-level accessibility indices. Utilizing both geographic proximity and capacity, we identify underserved areas to guide more effective resource allocation within the food bank network.

10 - Building Dynamic Capacity Function for Receiving Donations in Hunger-Relief Networks -Poster**Naimur Rahman Chowdhury, North Carolina State University, Raleigh, NC, United States**

Donation receiving capacity in a food bank network is influenced by the operational performance of food agencies, which changes over time. In this study, we perform dynamic agency categorization to build a capacity function leveraging deep learning and unsupervised learning. We demonstrate results with data from North Carolina.

11 - Optimizing Weight-Based and Nutritional Fairness Metrics in Food Bank Distribution - Poster

Nowshin Sharmile, North Carolina A&T State University, Raleigh, NC, United States

Food insecurity affects over 47 million people in the U.S. This study develops a lexicographic model to improve food bank distribution using the HER Nutrition Guidelines. The model balances weight-based fairness with nutrition-tier equity, considering varying supply levels and agency preferences to explore trade-offs between total pounds and nutritional quality.

12 - An Order-Based Bayesian Network to Analyze Diagnostic Delay and Type 2 Diabetes-Related Outcomes**Yukti Kathuria, University of Massachusetts Amherst, Amherst, MA, United States**

We developed a Bayesian Network model utilizing temporally ordered socio-demographic and healthcare utilization data to analyze the relationship between diagnostic delays and Type 2 Diabetes outcomes in primary care. Results indicated that higher glucose levels at diagnosis worsened long-term glycemic variability, and care continuity influenced outcomes based on diagnostic delay.

13 - Monotonic Mixture of Experts for Shape-Constrained Machine Learning**Yating Fang, Rutgers University, Piscataway, NJ, United States**

We introduce a shape-constrained Mixture-of-Experts (MoE) model that enforces global monotonicity—a property not guaranteed by monotonic experts alone. Our theoretical analysis proves this gap, and we propose a novel optimization framework that binds experts via shared constraints. The method is validated on tasks including molecular dynamics simulations of alloy failure, where it outperforms unconstrained counterparts.

14 - Assessment of Primary Healthcare Accessibility in Rural Areas - Poster**Valentina Castaneda-Torres, Georgia Institute of Technology, Atlanta, GA, United States**

Providing primary healthcare (PHC) is a global challenge, particularly in rural areas where nearly half of the population lives and lacks adequate access to essential healthcare services. This research addresses the assessment of accessibility to PHC in rural areas, where multiple socioeconomic, geographic, and infrastructure factors limit service accessibility and exacerbate populations' vulnerability. We propose a p-median facility location problem and explicitly model vulnerability by employing deprivation cost functions. We conduct a case study in a rural region in Colombia, and a series of sensitivity analyses of the impact of deprivation and terminal costs on PHC access. The findings show that 22% of the municipalities lacks access to PHC service points, and among those with access, 20% shared PHC service points and ranked highest in deprivation costs. This research also shows that terminal costs highly influence the optimal allocation of PHC service points to populations, conditioning the prioritization of more vulnerable populations when allocating essential resources.

15 - Design and Layout of a 100-Kw Scale Receiver Test Facility Using An Available Tower**Sebastian Dogue, Colorado School of Mines, Golden, CO, United States**

This study explores converting a 60-meter tower in Arvada, Colorado, into a 200-kWth concentrating solar test facility. The National Renewable Energy Laboratory's SolarPILOT was used to model heliostat field layouts and receiver positions to evaluate solar flux performance. Two layout strategies, dense and spaced, were compared to optimize efficiency, flux uniformity, and cost. The facility aims to support the cost-effective development of advanced heliostat and solar receiver technologies.

16 - A Scalable Deterministic Global Optimization Algorithm for Constraint Clustering Problems - Poster**Pedro Chumplitaz, University of South Florida, Tampa, FL, United States**

Constrained clustering leverages limited domain knowledge to improve clustering performance and interpretability, but incorporating pairwise must-link and cannot-link constraints is an NP-hard challenge, making global optimization intractable. Existing mixed-integer optimization methods are confined to small-scale datasets, limiting their utility. We propose Sample-Driven Constrained Group-Based Branch-and-Bound (SDC-GBB), a decomposable branch-and-bound framework that collapses must-linked samples into centroid-based pseudo-samples and prunes cannot-link constraints through geometric rules, while preserving convergence and guaranteeing global optimality. By integrating grouped-sample Lagrangian decomposition and geometric elimination rules for efficient lower and upper bounds, the algorithm attains scalability via embarrassingly simple parallelism. Experimental results show that our approach handles datasets with 200,000 samples for cannot-link constraints and 1,500,000 samples for must-link constraints—200–1500x larger than the current state of the art under comparable settings—while reaching an optimality gap of < 3 %. In providing deterministic global guarantees, our method also avoids the search failures that off-the-shelf heuristics often encounter on large datasets.

17 - Shortest Path: Ramp Optimization in Open Pit Mining**Felix Ayaburi, Colorado School of Mines, Lakewood, CO, United States**

We model haul roads as 3D graphs and use A* with slope- and turn-based penalties to optimize truck paths. Implemented in Python with NetworkX, our approach improves fuel efficiency for Komatsu 930Es and Hitachi EH5000ACs. Results show reduced fuel consumption. Future work includes subsurface planning and trade-off analysis for long-term efficiency.

18 - Multi-Objective Optimization for Equitable Electric Vehicle Charger Siting - Poster**Kevin Bamwiso, Colorado School of Mines, Lakewood, CO, United States, Estifanose Shiferaw**

As electric vehicle adoption accelerates, public Level 2 charging infrastructure must balance accessibility, usage, and cost. We present a tri-objective mixed-integer program that co-optimizes social equity, demand fulfillment, and grid-constrained costs when siting and sizing charger infrastructure. Case studies in Seattle and Pittsburgh demonstrate financially viable, socially inclusive expansion pathways.

19 - Predictive Maintenance for Steel Mill Roll Wear Using Large-Scale Data Analysis**Edikan Udofia, Colorado School of Mines, Lakewood, CO, United States**

Developing predictive models for steel mill roll wear using large industrial datasets (1GB+). Random Forest algorithms predict wear from steel grade, temperature, and coil properties. Campaign identification handles data quality challenges including roll reuse patterns. Achieving $R^2 > 0.85$ for maintenance optimization and cost reduction through data-driven scheduling.

20 - Integrated Learning and Optimization for Ramping-Cost Minimization In Real-TIME Electricity Markets - Poster**Imran Pervez, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia**

We developed a new integrated learning and optimization (ILO) methodology to predict context-aware unknown parameters in economic dispatch (ED), a crucial problem in power systems solved to generate optimal power dispatching decisions to serve consumer load. The ED formulation in the current study consists of load and renewable generation as unknown parameters in its constraints predicted using contextual information (e.g., prior load, temperature). The ILO framework train a neural network (NN) to estimate ED parameters by minimizing an application-specific regret function which is a difference between ground truth and NN-driven decisions favouring better ED decisions. We thoroughly analyze the feasible region of ED formulation to understand the impact of load and renewable learning together on the ED decisions. Corresponding to that we developed a new regret function to capture real-time electricity market operations where differences in predicted and true loads are corrected by ramping generators in real-time but at a higher cost than the market price. The proposed regret function when minimized using ILO framework train the NN to guide the load and renewable predictions to generate ED decisions favouring minimum generator ramping costs. This is unlike conventional sequential learning and optimization (SLO) framework which train NN to accurately estimate load and renewable instead of better ED decisions. The combined training of load and renewable using ILO is a new concept and lead to significantly improved ramping costs when compared with SLO based training of load and renewable and SLO trained load with 100% accurate renewable proving its decision-focused capability.

21 - A Dynamic Ensemble Method for Spatio-Temporal Wind Energy Forecasting - Poster**Jiaxiang Ji, Rutgers University, New Brunswick, NJ, United States**

The reliable operation of wind farms depends heavily on accurate wind energy forecasts across multiple look-ahead horizons. A central challenge is that methods optimized for short-term often perform poorly at longer horizons, and vice versa. To address this, we propose a dynamic ensemble framework that integrates two spatio-temporal models, each tailored to a specific forecast horizon. Applied to real-world data from the U.S. Mid-Atlantic region, our approach achieves considerable improvements.

22 - An Optimization Model for Urban Food Networks with Diverse Facilities and Access Constraints

Meen-Sung Kim, Georgia Institute of Technology, Atlanta, GA, United States

This study develops an optimization model to design urban food networks that improve access to fresh food for underserved populations. The model incorporates accessibility constraints, facility capacity, and minimum consumption thresholds, while accommodating diverse food outlet types. It explores trade-offs in placement decisions affecting proximity, coverage, and equity. The model is being applied in a Chicago-area case study to demonstrate real-world potential.

23 - Microbiome-Aided Classification of Forensic Body Fluid Mixtures via Conjecturing and Machine Learning

Samuel Paul Gyamfi, Virginia Commonwealth University, Henrico, VA, United States

We develop a microbiome-informed forensic pipeline that classifies biological samples as mixtures or single-source fluids using qPCR-derived Cq values for *S. salivarius*, *L. crispatus*, and *B. uniformis*. Several supervised models were employed to distinguish mixtures and infer their components, with top accuracies achieved by SVM, Random Forest, and Neural Networks. We then use invariant and property-based conjecturing to derive interpretable bounds between bacterial markers, aiding forensic explainability and enhancing microbial evidence modeling.

24 - Rethinking College Admissions: Can Test-Optional Policies Even the Playing Field?

Zijin Zhang, Carroll School of Management, Boston College, Chestnut Hill, MA, United States

This paper studies how test-required (TR) and test-optional (TO) policies affect academic merit and socioeconomic representation in college admissions. Using a game-theoretical model, we show that under TO, two forces—pool expansion and signal enhancement—jointly disadvantage middle-class students. Our results also suggest that the TO policy does not necessarily improve the representation of low- socioeconomic students at the cost of academic merit, challenging the popular belief. Empirical evidence from 3,701 U.S. colleges (2000–2019) supports these findings, revealing that TO often amplifies institutional biases and creates unequal outcomes.

SC01

Building A Level 3 A301

George B. Dantzig Dissertation Award

Award Session

George B Dantzig Dissertation Prize Session

Co-Chair: Daniel Freund, MIT, Cambridge, MA, United States

1 - Analytics for Better Urban Cycling

Bo Lin, University of Toronto, Toronto, ON, Canada

Cycling has become an increasingly popular transportation mode, offering significant benefits for urban mobility, public health, economic growth, and sustainability. However, safety and comfort concerns remain significant barriers to cycling adoption. Toward the vision of making biking safer and more inviting, this thesis develops quantitative methods for the efficient assessment, planning, and utilization of urban bike infrastructure, with a particular focus on fostering cycling adoption in Toronto, Canada. First, we introduce a computer vision approach to assess cycling stress—the discomfort cyclists experience on urban road networks. This method leverages the widespread availability of street-view images to replace current data- and labour-intensive assessment practices. Next, we formulate an optimization model to determine the optimal locations for new bike lanes, aiming to maximize the city's low-stress cycling accessibility—a metric strongly correlated with cycling mode choice in Toronto. Given the size and complexity of Toronto's road network, this optimization model cannot be solved by any existing methods. In response, we develop a machine learning-augmented optimization approach that is computationally efficient and produces provably high-quality solutions. Finally, we investigate cycling path recommendations in the context of last-mile delivery, where cycling couriers often deviate from delivery routes prescribed by platforms, leading to increased delivery times and inefficiencies in order batching and assignment, both of which depend heavily on accurate route modelling. To address this, we propose a data-driven approach to prescribing delivery paths that are both high-quality and perceived as such by couriers. The former property ensures service quality, while the latter promotes adoption.

2 - Understanding and Optimizing Societal Systems: Methods and Applications

Zhi Liu, Cornell Tech, Seattle, WA, United States

Modern societal systems are increasingly complex and present challenges in (1) understanding their impact through the vast amount of data collected, and (2) designing policies that promote efficient and equitable usage. In this dissertation, we describe methods and applications that tackle these challenges in two real-world systems: the urban crowdsourcing system and the urban library system. In doing so, we demonstrate how to leverage robust theoretical modeling and empirical analyses together, to uncover insights from data, and inform policy design.

3 - New Theory and New Practical Methods for Solving Large-Scale Linear and Conic Optimization

Zikai Xiong, MIT Operations Research Center, Cambridge, MA, United States

In the last several years there has been a dramatic shift in the way many large-scale linear programs (LPs) are solved in practice, with classic methods (etc, simplex method and interior-point method) being replaced by the primal-dual hybrid gradient method (PDHG) to solve large-scale LP problems. While PDHG has been very successful in solving large-scale LP problems, its performance can have substantial variance and an intuitive understanding of the drivers of its performance has been lacking. In this context, this dissertation (i) develops new theory to explain PDHG's performance in the worst case and the "average" case through deterministic and probabilistic analysis, (ii) develops new practical methods (central-path conic rescaling) for large-scale LPs based on the new theory, (iii) generalizes the new theory and methods to a broader class of conic optimization problems, and (iv) develops a practical GPU-based general-purpose conic program solver for solving large-scale problems including but not limited to quadratic programs, second-order cone programs and exponential cone programs.

4 - Models and Algorithms for Safeguarded Data-Driven Decision Making

Aras Selvi, Imperial College Business School, London, United Kingdom

This thesis focuses on making safeguarded data-driven decision making applicable and scalable in practice, while minimizing the cost of safeguarding. Privacy and fairness safeguards can reduce the utility gained from data, while robustness safeguards often lead to models that are intractable in real-world settings. The work in this thesis develops optimization frameworks for privacy-preserving mechanisms, applies fairness constraints in sequential decisions such as housing allocation with LAHSA, and revises robust models to handle categorical and synthetic data in collaboration with JP Morgan.

SC02

Building A Level 3 A302

IISE Transactions: Insights and Perspectives from the Editorial Board

Panel Session

Committee's Choice

Co-Chair: Judy Jin, University of Michigan, Ann Arbor, MI, United States

1 - Moderator Panelist

Judy Jin, University of Michigan, Ann Arbor, MI, United States

This panel session aims to provide valuable insights into IISE Transactions, the flagship journal of the Institute of Industrial and Systems Engineers (IISE). The session will cover the journal's overall scope and clarify distinctions among its four Focused Issues: Data Science, Quality, and Reliability (DSQR); Design and Manufacturing (DM); Operations Engineering and Analytics (OEA); and Supply Chain and Logistics (SCL). The panelists, including the Editor-in-Chief and the four Focused Issue Editors, will discuss the peer-review process, highlight emerging research trends, and offer practical guidance for preparing successful submissions. Given the journal's broad relevance and high-impact contributions across multiple domains, this is a community choice session, led by the Quality, Statistics, and Reliability (QSR) Section

and co-sponsored by the Data Mining (DM) Section, INFORMS Computing Society (ICS), the Society on Location Analysis (SOLA), and the Transportation Science and Logistics (TSL) Section.

2 - Panelist

Kaibo Liu, UW-Madison, Madison, WI, United States

3 - Panelist

Zhenyu Kong, Virginia Tech, Blacksburg, VA, United States

4 - Panelist

Siqian Shen, University of Michigan, Ann Arbor, MI, United States

5 - Panelist

Jennifer Ryan, Lehigh University, Bethlehem, PA, United States

SC03

Building A Level 3 A303

The MCDM Junior Researcher Best Paper Award Session

Award Session

Multi Criteria Decision Making

Chair: Salvatore Greco, University of Catania, Catania, Italy

Co-Chair: Banu Lokman, University of Portsmouth, Portsmouth, United Kingdom

1 - A Novel Spatiotemporal Multi-Attribute Method for Assessing Flood Risks in Urban Spaces Under Climate Change and Demographic Scenarios

Lucas da Silva, Universidade Federal de Pernambuco (UFPE), Recife, Brazil, Marcelo Alencar, Adiel de Almeida

Addressing the challenge of building safer societies in the context of climate-induced extreme events is complex, particularly due to the interplay of uncertain and sometimes contradictory urban policies. Given this backdrop, this research proposes a novel decision-making framework grounded in non-stationary multi-attribute utility theory (NSMAUT), which incorporates human, environmental, and economic dimensions to evaluate flood risks in urban settings. Building upon expected utility theory and extending the multi-attribute utility theory (MAUT), this approach represents a methodological advancement by addressing dynamic preferences over time. A key feature of the model is its ability to account for temporal preferences, recognizing that decision-makers may demonstrate greater patience when considering outcomes that occur further in the future, specially climate and demographic scenarios.

This time-sensitive perspective is essential in urban adaptation planning in a changing climate, where the advantages of resilience measures often unfold gradually. The model is applied in a numerical case study, supported by climate and demographic modeling, statistical tools and simulation techniques, to thoroughly explore its practical use. Findings suggest that this approach can significantly improve the formulation of urban adaptation policies by offering a dynamic and multidimensional tool for assessing climate-related risks. By reflecting real-world complexities such as uncertainty and time-delayed impacts, the model supports the creation of more adaptive and robust strategies for urban resilience in an era of intensifying climate threats.

2 - Towards Explainable Interactive Multiobjective Optimization: R-XIMO

Giovanni Misitano, University of Jyväskylä, Jyväskylä, Finland, Bekir Afsar, Giomara Lárraga, Kaisa Miettinen

Interactive multiobjective optimization methods guide decision makers toward preferred solutions in problems involving multiple conflicting objectives. By incorporating human preferences into the solution process, these methods support exploration of trade-offs and help identify satisfactory compromises. However, it is often unclear how the stated preferences influence the resulting solutions. This lack of transparency can reduce trust, hinder learning, and weaken the overall impact of the decision-support process.

To address this challenge, we propose a novel paradigm: explainable interactive multiobjective optimization. We introduce R-XIMO, a new method that integrates explainability into reference point-based interactive multiobjective optimization using tools from explainable artificial intelligence. Specifically, R-XIMO applies SHAP (Shapley Additive exPlanations) values to quantify and communicate how preferences shape the optimization outcomes.

R-XIMO enhances decision support in two ways. First, it helps the decision maker understand how their preferences affect the generated solutions and what trade-offs are involved. Second, it offers guidance for revising preferences, suggesting which objective to compromise to improve another of higher interest. This kind of feedback is lacking in existing interactive methods.

This research is one of the first to combine explainable artificial intelligence with interactive multiobjective optimization. By bridging these two fields, R-XIMO delivers explanations and strengthens the decision-support process. We demonstrate the method's potential through a numerical example and a case study.

Our findings suggest that incorporating explainability into interactive optimization holds significant promise. It enables more transparent, and effective decision-support, opening a new and exciting research direction at the intersection of optimization and explainable artificial intelligence.

SC04

Building A Level 3 A304

Reinforcement Learning in Finance

Invited Session

Finance

Chair: Nan Chen, The Chinese University of Hong Kong, 709A William Mong Engineering Building, Hong Kong, N/A

1 - Reconciling Discrete-Time Mixed Policies and Continuous-Time Relaxed Controls in Reinforcement Learning and Stochastic Control

Mathieu Lauriere, NYU Shanghai, Shanghai, China, People's Republic of, Rene Carmona

Reinforcement learning (RL) is currently one of the most popular methods, with breakthrough results in a variety of fields. The framework relies on the concept of Markov decision process (MDP), which corresponds to a discrete-time optimal control problem. In the RL literature, such problems are usually formulated with mixed policies, from which a random action is sampled at each time step. Recently, the optimal control community has studied continuous-time versions of RL algorithms, replacing MDPs with mixed policies by continuous-time stochastic processes with relaxed controls. In this talk, we rigorously connect the two problems: we prove the strong convergence of the former towards the latter when the time discretization goes to 0. Joint work with Rene Carmona (Princeton University).

2 - Algorithmic Collusion and Entry Deterrence Through Limit Pricing

Patrick Chang, Oxford University, Oxford, United Kingdom

We derive a theoretical model of oligopolistic behavior competition in market-making in a model with payment for order flow (PFOF). In a model with high-speed trading and no value from price discovery, a retail liquidity program (RLP) based on PFOF is a more efficient mechanism to provide liquidity as it allows liquidity providers to

avoid sniping from faster traders. Some RLP implicitly acknowledge this negative effect by requiring firms participating in RLPs to provide liquidity in the exchange. We find that such arrangements can backfire, leading to collusive equilibria, where the faster market-makers undercut slower market-makers in the exchange in order to exclude them and reduce competition for retail orders through the RLP. This undercutting can be interpreted as dumping liquidity in the exchange in order to corner the RLP.

3 - Mf-Oml: Online Mean-Field Reinforcement Learning with Occupation Measures for Large Population Games

Anran Hu, Columbia University, IEOR, New York, NY, United States

Reinforcement learning for multi-agent games has attracted significant attention in recent years. However, due to the complexity of solving Nash equilibria in large-population games, existing works with guaranteed polynomial complexity primarily focus on variants of zero-sum and potential games or aim to compute (coarse) correlated equilibria, particularly in online settings. In this talk, we propose MF-OML (Mean-Field Occupation-Measure Learning), an online mean-field reinforcement learning algorithm for approximating Nash equilibria in large-population sequential symmetric games. Evaluated in terms of cumulative deviation from Nash equilibria, the algorithm achieves a high-probability regret bound for monotone games. The regret bound consists of two components: one capturing the effect of mean-field approximation, which vanishes as the number of players approaches infinity, and the other that grows sublinearly with the number of episodes. As a by-product, we also establish the first tractable, globally convergent computational algorithm for solving Nash equilibria in monotone mean-field games.

4 - Multi-Task Dynamic Pricing in Credit Market with Contextual Information

Jingwei Ji, University of Southern California, Los Angeles, CA, United States, Adel Javanmard, Renyuan Xu

We study the dynamic pricing problem faced by a broker seeking to learn prices for a large number of credit market securities, such as corporate bonds, government bonds, loans, and other credit-related securities. A major challenge in pricing these securities stems from their infrequent trading and the lack of transparency in over-the-counter (OTC) markets, which leads to insufficient data for individual pricing. Nevertheless, many securities share structural similarities that can be exploited. Moreover, brokers often place small "probing" orders to infer competitors' pricing behavior. Leveraging these insights, we propose a multi-task dynamic pricing framework that leverages the shared structure across securities to enhance pricing accuracy. In the OTC market, a broker wins a quote by offering a more competitive price than rivals. The broker's goal is to learn winning prices while minimizing expected regret against a clairvoyant benchmark. We model each security using a d -dimensional feature vector and assume a linear contextual model for the competitor's pricing of the yield, with parameters unknown a priori. We propose the Two-Stage Multi-Task (TSMT) algorithm: first, an unregularized MLE over pooled data to obtain a coarse parameter estimate; second, a regularized MLE on individual securities to refine the parameters. We show that the TSMT achieves a regret bounded by $\tilde{O}(\sqrt{M d}) + M d$, outperforming both fully individual and fully pooled baselines, where M is the number of securities and d quantifies their heterogeneity.

5 - A Two Timescale Evolutionary Game Approach to Multi-Agent Learning and Its Application in Algorithmic Collusion Study

Mingyue Zhong, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Nan Chen

In this paper, we propose a novel two-timescale evolutionary game approach for solving general-sum multi-agent reinforcement learning (MARL) problems. Unlike existing literature that requires solving Nash equilibria exactly or approximately in each learning episode, our new approach synthesizes three key design components. First, we introduce a simple perturbed best response-based protocol for policy updates, avoiding the computationally expensive task of finding exact equilibria at each step. Second, agents use fictitious play to update their beliefs about other agents' policies, relaxing the requirement for observable Q-values of all agents as in classical Nash Q-learning. Third, our algorithm updates policies, beliefs, and Q-values at two different timescales to address non-stationarity during learning. The new approach provably converges to ϵ -Nash equilibria of MARL problems without imposing the global optima or saddle point conditions, two restrictive assumptions typically needed in the existing literature. AI-powered algorithms are increasingly used in marketplaces for pricing goods and services. The numerical experiments in this paper show that the realization of collusive Nash equilibrium is highly path-dependent and hinges on many factors, such as agents' impatience and the algorithm's exploration degree. Moreover, we also find that intervention timing is crucial for the effectiveness of regulatory policies.

SC05

Building A Level 3 A305

JFIG Paper Competition I

Award Session

Junior Faculty Interest Group

Chair: Karmel Shehadeh, University of Southern California, Los Angeles, CA, United States

Co-Chair: Raed Al Kontar, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Ozlem Ergun, Northeastern University, Newton, MA, United States

1 - Social Learning with Limited Attention: Negative Reviews Persist under Newest First

Jackie Baek, New York University, New York City, NY, United States, Atanas Dinev, Thodoris Lykouris

We study a model of social learning from reviews where customers are computationally limited and make purchases based on reading only the first few reviews displayed by the platform. Under this limited attention, we establish that the review ordering policy can have a significant impact. In particular, the popular Newest First ordering induces a negative review to persist as the most recent review longer than a positive review. This phenomenon, which we term the Cost of Newest First, can make the long-term revenue unboundedly lower than a counterpart where reviews are exogenously drawn for each customer.

We show that the impact of the Cost of Newest First can be mitigated under dynamic pricing, which allows the price to depend on the set of displayed reviews. Under the optimal dynamic pricing policy, the revenue loss is at most a factor of 2. On the way, we identify a structural property for this optimal dynamic pricing: the prices should ensure that the probability of a purchase is always the same, regardless of the state of reviews. We also consider a setting where product quality evolves over time according to a Markov chain; we find that Newest First better tracks current quality but still leads to lower revenue, highlighting a trade-off between customer belief accuracy and revenue. Finally, we support our theoretical findings with numerical simulations and an empirical analysis on reviews from Tripadvisor.

2 - Can Price Discrimination be Progressive?

Max Biggs, University of Virginia, Charlottesville, VA, United States, Titing Cui, Michael Hamilton, Enfeng Xing

Traditional welfare analysis evaluates pricing policies based on aggregate revenue and customer surplus. However, modern approaches increasingly emphasize equity by asking not just how much aggregate surplus is derived, but by whom. In this work, we examine the impacts of price discrimination on market outcomes under a new notion of proportional surplus, which measures the percentage of a customer's valuation that is retained as surplus. We show proportional surplus provides a more equitable lens by which to view market outcomes, assigning greater weight to utility gained by customers with the least ability to pay. Further, we demonstrate that although traditional customer surplus often decreases when firms price discriminate, proportional surplus increases under moderate levels of revenue-maximizing price discrimination in many natural markets. These findings challenge the conventional belief that price discrimination harms customers and suggests instead that modest levels of price discrimination can improve outcomes for both firms and the markets that they serve when equity is considered.

3 - A Data-driven Approach to Improve Artisans' Productivity in Distributed Supply Chains

Divya Singhvi, New York University, New York, NY, United States, Somya Singhvi, Xinyu Zhang

Despite their vital role in the global rural economy, and as a major source of employment for women in the developing world, artisanal supply chains continue to be plagued by low productivity and high poverty levels. Identifying effective and implementable solutions to improve artisan productivity is a challenging task due to high fragmentation in the upstream parts of the supply chain. This paper presents research conducted in close collaboration with one of the leading exporters of handmade rugs in India. Leveraging insights from the field visits and detailed supply chain data, we provide robust empirical evidence that frequent supervisor visits can play a crucial role in improving artisans' productivity. Our results from Instrumental Variables analysis indicate that a one-day decrease in the average number of days between supervisor visits to remote weavers can increase weaving rates by 8.5%. We also find that (i) visits to looms with difficult-to-weave rugs and (ii) visits that are

consistently scheduled are associated with a more substantial positive impact on weavers' productivity. To capitalize on these insights, we propose an optimization framework for scheduling supervisor visits in the supply chain. Finally, we collaborate with our partner to implement this scheduling framework for a set of treatment looms over 25 consecutive weeks. A difference-in-differences (DID) analysis using data from almost 6,000 supervisor visits across 200 looms during the implementation suggests that the implementation is associated with a statistically significant 16.7% increase in the weaving rate for treatment looms, highlighting the practical relevance of the work.

4 - Strong Formulations and Algorithms for Regularized A-optimal Design

Yongchun Li, Chinese University of Hong Kong, Shenzhen, Shenzhen, Hong Kong

We study the Regularized A-Optimal Design (RAOD) problem, which selects a subset of k experiments to minimize the inverse of the Fisher information matrix, regularized with a scaled identity matrix. RAOD has broad applications in Bayesian experimental design, sensor placement, and cold-start recommendation. We prove its NP-hardness via a reduction from the independent set problem. By leveraging convex envelope techniques, we propose a new convex integer programming formulation for RAOD, whose continuous relaxation dominates those of existing formulations. More importantly, we demonstrate that our continuous relaxation achieves bounded optimality gaps for all k , whereas previous relaxations may suffer from unbounded gaps. This new formulation enables the development of an exact cutting-plane algorithm with superior efficiency, especially in high-dimensional and small- k scenarios. We also investigate scalable forward and backward greedy approximation algorithms for solving RAOD, each with provable performance guarantees for different k ranges. Finally, our numerical results on synthetic and real data confirm the efficacy of the proposed algorithms. We further showcase the practical effectiveness of RAOD by applying it to a real-world user cold-start recommendation problem.

SC07

Building A Level 3 A307

Smart Transportation Systems

Job Market Showcase

Job Market Showcase

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

1 - Optimizing Expeditionary Logistics: Dynamic Discretization for Fleet Management

Haden Boone, Georgia Institute of Technology, Atlanta, GA, United States, Mathieu Dahan

This paper introduces the Expeditionary Logistics Network Design Problem (ELNDP), a novel formulation that captures the operational complexity of military logistics in expeditionary environments, where unmet demand penalties dominate transportation costs and multi-modal fleet management is essential. We propose Solve-ELNDP, an iterative refinement algorithm based on dynamic discretization discovery (DDD), which extends prior work by incorporating asset management and multi-commodity flow optimization. Our method constructs partially time-expanded networks that adaptively refine temporal resolution, enabling high-fidelity modeling without the computational burden of full time expansion. Theoretical results establish the validity of our relaxation. Through a case study involving 68 realistic instances validated by the United States Marine Corps, Solve-ELNDP is able to solve 36.7% more instances, demonstrates up to a 56.4% reduction in solve times, and up to a 148.9% increase in demand fulfillment compared to traditional formulations. These results highlight the potential of dynamic discretization discovery to support scalable, data-driven decision-making in high-stakes logistics planning.

2 - Optimal Control of Transport Systems: Leveraging Set-Based Methods and Complex Analysis

Zhexian Li, University of Southern California, Los Angeles, CA, United States

Transport behaviors are fundamental to the effective operation of civil infrastructure systems, such as urban transportation networks and water distribution systems. Optimally controlling these behaviors, typically modeled by partial differential equations (PDEs), significantly enhances infrastructure performance and societal well-being. We investigate optimal control strategies for two classes of transport behaviors -- roadway traffic flow governed by nonlinear continuity equations, and solute transport in fluids described by linear evolution equations.

The first part addresses finite-dimensional traffic control problems, designing set-membership predictive control (Set-PC) methods for freeway ramp metering and intersection signal timing with partial traffic information. Using discretized traffic dynamics (Cell Transmission Model and deterministic queuing models), the proposed Set-PC framework maximizes network throughput, robustly handles incomplete information, and ensures bounded queue lengths.

The second part explores infinite-dimensional optimal control problems, presenting analytical and numerical solutions for optimal control of linear evolution equations with boundaries. Applying the unified transform (Fokas method) with complex analysis, a closed-form expression for the linear quadratic regulator of linear evolution equations is derived, revealing structural insights into optimal control. Additionally, numerical methods for optimal velocity control in advection-diffusion-reaction systems with state constraints are developed, demonstrating practical computational strategies for these complex PDE-driven systems.

3 - A Double Decomposition Algorithm for Network Planning and Operations in Deviated Fixed-Route Microtransit

Alexandria Schmid, MIT, Cambridge, MA, United States, Bernardo Martin-Iradi, Kayla Cummings, Alexandre Jacquillat

Microtransit offers opportunities to enhance urban mobility by combining the reliability of public transit and the flexibility of ride-sharing. This paper optimizes the design and operations of a deviated fixed-route microtransit system that relies on reference lines but can deviate on demand in response to passenger requests. We formulate a Microtransit Network Design (MiND) model via two-stage stochastic integer optimization, with a first-stage network design and service scheduling structure and a second-stage vehicle routing structure. We derive a tight second-stage relaxation using a subpath-based representation of microtransit operations in a load-expanded network. We develop a double-decomposition algorithm combining Benders decomposition and subpath-based column generation. We prove that the algorithm maintains a valid optimality gap and converges to an optimal solution in a finite number of iterations. Results obtained with real-world data from Manhattan show that the methodology scales to large and otherwise-intractable instances, with up to 10-100 candidate lines and hundreds of stops. Comparisons with transit and ride-sharing suggest that microtransit can provide win-win outcomes toward efficient mobility (high demand coverage, low costs, high level of service), equitable mobility (broad geographic reach) and sustainable mobility (limited environmental footprint). We provide an open-source implementation to enable replication.

4 - Optimizing Freight Electrification: Stochastic Routing and Grid-Aware Infrastructure Design

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

Electrifying heavy-duty trucks is critical for reducing emissions in logistics but remains challenging due to high energy demands, operational uncertainties, and infrastructure limitations. This talk presents innovative strategies to overcome these challenges, including advanced optimization for routing and infrastructure design. I will show our methods of energy-efficient routing for mixed fleets of electric and conventional trucks under energy consumption uncertainty. Reformulated as a second-order cone mixed-integer program, this approach reduces operating costs, cuts emissions, and greatly reduces route capacity violations under real-world energy consumption uncertainty. In addition, I will compare battery swapping and fast charging mechanisms for heavy-duty freight from different perspectives. Focusing on California's logistics network, we introduce an optimization framework for station location, sizing, and battery sizing, leveraging real-world data on freight traffic, energy prices, renewable generation, and grid services. Through evaluations of transportation efficiency, battery utilization efficiency, and grid service potential, we provide pioneering insights into charging infrastructure networks for sustainable freight systems. Finally, I will discuss open challenges in operation and grid-aware infrastructure planning, and mobility operation, how scalable solutions are desired in this area, and my proposals to address them.

SC08

Building A Level 3 A310

GenAI and User Productivity

Invited Session

Information Systems

Chair: Lei Wang, Indiana University, Bloomington, IN, 47405, United States

1 - AI-Powered Digital Streamer in Live Streaming

Si Xie, George Mason University, Fairfax, VA, United States

Livestream shopping has emerged as a popular method for promoting and selling products online, traditionally featuring human livestreamers. However, with advancements in generative artificial intelligence (GenAI), virtual livestreamers (VLS) capable of natural language processing and gesturing have been introduced to assist human hosts.

In this study, we analyze livestream session-level data from over 500 livestreaming channels on a major Chinese livestream shopping platform to assess the impact of VLS adoption on sales. Our findings indicate that while integrating virtual livestreamers initially boosts session sales, this positive effect diminishes over time. Specifically, we find that VLS performance improves in responding to viewers' questions, yet viewers spend less time watching the sessions, ultimately leading to reduced sales. This paradox likely stems from the increased efficiency of VLS, which enables viewers to obtain satisfactory answers more quickly, spending less time watching more product presentations.

To mitigate this decline, we explore potential strategies and find that using non-humanoid, engaging VLS interfaces—such as a mascot—helps sustain sales performance. Our results highlight the importance of carefully implementing generative AI to maximize its long-term effectiveness in driving sales.

2 - Transparent-by-Design: a Retrieval-Augmented Information System for Interpretable Healthcare Predictions

Sizhe(Jasmine) Chen, University of Utah, Salt Lake City, UT, United States, Paul Hu, Da Xu

While artificial intelligence (AI) offers immense potential for healthcare, its clinical adoption is hindered by interpretability challenges: most predictive models operate as opaque "black boxes," existing eXplainable AI (XAI) methods fail to produce narrative evidence essential for clinical reasoning, and generating such explanations raises data privacy concerns when deploying large language models in healthcare settings. To address these issues, this study proposes a novel Retrieval-Augmented Generation (RAG) framework that bridges the gap between AI systems and clinical practice. Our method leverages a multi-agent system to perform knowledge-guided retrieval of similar patient cases from institutional databases, generating case-based evidence for its predictions through local inference deployment that ensures complete data privacy compliance. To ensure rigor, we conduct a multi-faceted evaluation. We assess predictive performance using standard clinical metrics (AUC, F1-Score) and validate the explanations of our RAG model through conceptual evaluation property (Co-12) framework and a small-scale user study with healthcare professionals. Our RAG-based approach transforms opaque predictions into trustworthy, evidence-based decision support tools. By providing interpretable predictions that respect data privacy, our work facilitates the deeper integration of AI into clinical workflows, enabling clinicians to derive greater value from health information systems.

3 - A Domain-Adversarial, Two-Stage Ensemble Learning Method for Predicting the Length of Stay of Trauma Patients

Jiaxuan Peng, The University of Utah, Salt Lake City, UT, United States, Da Xu, Paul Hu, Tingshuo Huang

Length of stay (LOS) management is crucial to resource allocation efficiency and patient outcome, especially for trauma patients who often need hospitalization care. Data-driven analytics supports LOS predictions but are constrained the number of trauma patients in healthcare organizations and the highly skewed LOS distributions among patients. We address these challenges and develop a domain-adversarial, two-stage ensemble learning method that leverages the large-scale National Trauma Data Bank (n = 352,391) and a proprietary patient cohort in a healthcare organization in Taiwan (n = 9,485). This method incorporates a domain -adversarial feature extractor that aligns the heterogeneous feature distributions between the two datasets to learn shared domain-invariant representations and classify each patient as of short (< 7 days) or long (≥ 7 days) LOS. The shared representations, together with the classification probabilities, then inform regressors to derive initial domain-invariant LOS estimates (in days). In this (second) stage, a feedforward neural network, trained on domain-specific features learned from the proprietary patient data, generates (target) domain-specific LOS estimates. Finally, ensemble learning is applied to fuse both domain-invariant and domain-specific predictions and produce a LOS estimate for each patient. Evaluations show our method's mean absolute error is significantly lower than those of prevalent methods that represent different analytics approaches, thus demonstrating the value of adversarial domain adaptation, coupled with a classification-driven regression cascade and ensemble learning, for estimating patients' LOS and suggests a scalable strategy for harnessing large data registries to augment analyses of proprietary patient data for downstream predictions.

4 - Is Algorithm Enough? An Informational Perspective on Viewer Targeting in Live-Streaming Commerce

Lei Wang, Indiana University, Bloomington, IN, United States, Qingping Lin, Jing Tian, Ning Jia

In order to attract traffic in the live streaming, many merchants are increasingly implementing the algorithmic viewer targeting (AVT) strategy. After viewers acquisition through the AVT strategy, interactive engagement subsequently assumes the central role in bridging the gap between traffic and conversion. Thus, this strategy requires close coordination with engagement-based information to retain traffic and drive conversions. To assess the effectiveness of the AVT strategy, we compare the responses of targeted viewers to organic viewers regarding two key sources of interactive information: streamer expressions, including both informative and persuasive content, and viewer comments, comprising both self-generated and other-generated comments. Specifically, we also consider the influence of two different types of streamers: influencers and non-influencers. Our findings contribute to the growing literature on the effectiveness and mechanisms of AVT in live streaming and offer practical insights for optimizing information provision strategies under AVT across different types of livestreams.

SC09

Building A Level 3 A311

Information Disclosure

Job Market Showcase

Job Market Showcase

Chair: YANG XIA, City University of Hong Kong, Hong Kong, Hong Kong

1 - Information Disclosure Policies for Omnichannel Service with Invisible Customers

Guanling Yang, Simon Business School, University of Rochester, Rochester, NY, United States, Ricky Roet-Green

The omnichannel service, which integrates brick-and-mortar stores and online channels, has become the prevailing standard in the service industry. While walk-in customers are physically in the store and therefore visible to others, online customers are typically invisible due to the inherent nature of the online channel. This invisibility poses a significant challenge for customers' decisions about whether to wait for service, and service providers' ability to manage the service system. In our study, we investigate different information-sharing policies that reflects common real-world practices. We find that information invisibility can, in many cases, increase revenue while reducing social welfare. Perhaps surprisingly, in certain market segments, information invisibility can enhance social welfare and customer satisfaction. These findings underscore the importance of tailoring information disclosure policies to the specific characteristics of each service segments. In particular, we show that the optimal policy depends on the prevalence of online orders and the normalized service value perceived by customers.

2 - The Value of Community Information for Pricing Under Network Externalities

Calvin Roth, University of Minnesota, Minneapolis, MN, United States, Ankur Mani, Jiali Huang

We study the value of community information for pricing decisions for a merchant selling divisible goods with positive network externalities to consumers in a social network. In this setting, we consider consumers whose utility functions are homogeneous except for their position in the network. We study optimal pricing structures when there is uncertainty about the network structure and provide general expressions for the optimal price and profit. Prior work has shown that in a large class of networks, pricing using the information about the social network does not improve profit significantly. Yet, it is unclear whether there is value of network information for some random networks. We study the value of network information for dense stochastic block models. For these networks, each block gives information on how communities on average interact with the rest of the network instead of consumers. We find that for this class of networks, there will be significant losses when pricing decisions are made with no information about the network compared to optimal profits under full network information. We also find that for asymptotically large networks making pricing decisions using only the information about community structure and community memberships of the consumers does not introduce any significant change in expected profits over optimal pricing under full network information. Further, we show that as the community information becomes noisy, coarser price structure provides higher profit and when the noise density exceeds the structured network density then optimal price is uniform and network information has no value.

3 - Patenting and Information Disclosure

Xizhao Wang, Kellogg School of Management, Northwestern University, Evanston, IL, United States

Invention disclosure facilitates knowledge spillovers, supporting future progress but potentially limiting appropriability for the inventor. This paper examines invention disclosure behavior by analyzing the readability of patent texts, using both traditional and novel AI-based readability scores. Using two difference-in-differences analyses, the study finds that following the 1980 Bayh-Dole Act and the establishment of Technology Transfer Offices, university-affiliated inventors reduced the readability of patent detailed descriptions. This decrease in readability does not extend to patent summary texts, suggesting that university inventors strategically limit information on how to make and use the invention. The findings reveal strategic disclosure behavior not just in the decision of whether to patent or keep inventions as trade secrets, but also in the degree of patent language clarity. Institutional changes lead inventors to selectively adjust the information disclosed in their patents. The mechanisms behind this influence and its impact on follow-on innovation are further explored.

4 - Efficient Queueing via Partial Bayesian Persuasion: Nudging Just Enough Customers Amid Quality Uncertainty

YANG XIA, City University of Hong Kong, Hong Kong, Hong Kong, Pengfei Guo, Qiao-chu He, Yulan Wang

We study the information design problem for a service provider whose service quality fluctuates between high and low levels, and only a portion of customers receive information about its service quality. We model this as a Bayesian persuasion problem, where the service provider sends a binary quality signal to a portion of customers based on a committed information strategy after the service quality level is realized. Customers then update their beliefs using Bayes' rule and decide whether or not to queue. We analyze two cases: one in which the persuasion coverage level is fixed, and the other in which it is a decision variable. We derive the optimal information design strategies for both profit maximization and welfare maximization. When the persuasion coverage level is fixed, the profit-maximizing provider often adopts a partial disclosure strategy: sending a positive signal when service quality is high and randomizing between positive and negative signals otherwise. By contrast, the welfare-maximizing provider's information strategy depends on market size: full disclosure in small-sized markets, where each signal fully reveals the service quality; partial disclosure in medium-sized markets, where the service provider randomizes between positive and negative signals when service quality is high and sends a negative signal otherwise; and non-disclosure in large-sized markets, in which it sends the same signal regardless of the realized quality level. When the persuasion coverage level is a decision variable, we show that irrespective of the provider's objective, partial persuasion always outperforms full persuasion when the market size is sufficiently large.

SC10

Building A Level 3 A312

Optimization in Energy and Sustainability

Job Market Showcase

Job Market Showcase

Chair: Austin Saragih, Massachusetts Institute of Technology, Cambridge, MA, 02142, United States

1 - Spin the Bottle Bill: Deposit-Refund System Policy and Reverse Supply Chain Design

Austin Saragih, Massachusetts Institute of Technology, Cambridge, MA, United States, Lara Pontes, Saurabh Amin, Jan Fransoo

The state of California's Bottle Bill, a deposit-refund system policy, charges consumers a deposit upon beverage purchase and refunds it upon return of the empty container to a recycling center. In California and many other states, the refund value has remained stagnant at 5 cents for decades, and the number of recycling centers has decreased, leading to a decline in recycling rates. To address these issues, we spin the Bottle Bill by jointly optimizing its policy and reverse supply chain as a bi-level mixed-integer nonlinear program. At the policy design level, we optimize the redemption value, manufacturers' fees, and distance to recycling centers. Given this distance policy, at the network design level, we then optimize the number of recycling centers for each county in the state. To solve the bi-level problem, we devise an approach that transforms it into a single-level problem using continuous approximation, and solves it to global optimality using monotonicity analysis. Applying our approach to California's Bottle Bill, we provide policy options and insights that the state could implement to achieve its 80% recycling goal. These options and insights can also be adapted to other states and countries to advance the circular economy.

2 - Multi-Objective Optimization for Sustainable Watershed Management: Algorithms, Tools, and Decision Support System

Ashim Khanal, University of South Florida, Tampa, FL, United States, Hadi Charkhgard

We introduce a suite of advanced optimization algorithms and automated decision-support tools designed for nutrient management in large-scale watershed systems. Central to this work is AquaNutriOpt, a Python-based framework that formulates and solves large-scale mixed-integer network optimization models under a multi-objective setting. These models determine cost-effective combinations of Best Management Practices (BMPs) and Treatment Technologies (TTs) to minimize phosphorus and nitrogen loads, subject to budgetary and land-use constraints. The framework supports bi-objective optimization over multiple time periods, accounting for the variable nature of nutrient loading and enabling decision-makers to explore robust trade-offs. In addition to the traditional approach of approximating the entire trade-off frontier between conflicting objectives, we propose a complementary strategy that directly balances these objectives using Nash social welfare optimization, which produces more equitable solutions. However, because exact methods are often computationally expensive, we introduce a novel feasibility pump heuristic in the criterion space to solve these problems efficiently. In the context of controlling harmful algal blooms (HABs), this approach maximizes the product of phosphorus and nitrogen reduction loads, thereby directly targeting reductions in chlorophyll-a levels—a key indicator of HABs that is strongly influenced by the phosphorus-to-nitrogen (P:N) ratio. We further present a web-based decision support system that automates hydrological model selection with user inputs, optimization execution, and output visualization through an intuitive interface. Input data automation for popular models such as WAM and SWAT, and custom data for other models is available. Collectively, these tools provide scalable, high-performance, and actionable solutions for sustainable watershed management.

3 - Navigating Ambiguity in Scheduling: Distributionally Robust Workload Coordination in Data Centers

Yifu Ding, Massachusetts Institute of Technology, Cambridge, MA, United States, Thomas Magnanti

Amid rising energy demands from data centers (DCs), cloud providers seek solutions to reduce power consumption and alleviate grid stress. Dynamic Voltage and Frequency Scaling (DVFS) offers a promising approach that adjusts CPU/GPU power usage to enable demand response. However, the effectiveness of DVFS is impacted by the volatility and uncertainty in real-time CPU utilization, which can lead to rebound effects in the absence of robust scheduling mechanisms. The scheduling problem is critical in many domains, such as transportation and computer system management. This work models a group of servers coordinated by a market aggregator scheduling for grid demand reduction. To address ambiguity in CPU utilization, we propose a scenario-based distributionally robust optimization (DRO) formulation that captures the dynamics of power consumption over multiple time stages. Our approach is evaluated using synthetic and real-world datasets within a scalable power market game framework.

4 - Decentralized Der Integration into Wholesale Energy Markets via Reinforcement Learning of Mean-Field Games

Jun He, Purdue University, West Lafayette, IN, United States, Andrew Liu

FERC Order 2222 facilitates the integration of distributed energy resources (DERs) into wholesale energy markets, but specific mechanisms are needed for effective prosumer participation through aggregators. Current research primarily focuses on how a single aggregator manages DER portfolios, often under the simplifying assumption that wholesale market prices, such as locational marginal prices (LMPs), are exogenously fixed. While this perspective provides valuable insights into aggregator operations, it overlooks the feedback loop where aggregator actions influence market outcomes.

In response, we propose a hybrid Mean-Field Control (MFC) and Mean-Field Game (MFG) framework for integrating DER aggregators into these markets with a large number of agents. Unlike traditional approaches, our model captures the feedback between aggregators' strategies and LMPs while enabling decentralized decision-making. This algorithm enables aggregators to maintain beliefs about LMPs at a long-run equilibrium, solve their own Markov decision problems, and determine prosumers' supply or demand bids, which are then aggregated for the wholesale market. System operators continue running multi-settlement systems with economic dispatch. We prove the existence of a mean-field equilibrium (MFE) and provide a reinforcement-learning-based algorithm for prosumers to learn to converge to an MFE with entropy regularization. Numerical results show that LMPs can quickly reach a steady state. Furthermore, numerical comparisons with and without energy storage illustrate that our model can prevent extreme LMP values, fostering a more stable market even with completely decentralized decision-making.

SC11

Building A Level 3 A313

Applications in Public Health

Job Market Showcase

Job Market Showcase

Chair: Yuming Sun, Georgia Institute of Technology, Atlanta, GA, United States

1 - When More Medical Decisions Mean Better Healthcare

Suyanpeng Zhang, Northwestern University, Evanston, IL, United States

Frequent and data-driven medical decisions can significantly enhance patient care. This talk explores how increasing decision-making opportunities leads to improved outcomes. We present methodological advances in Markov models with applications in disease management.

2 - Optimizing Access to Vaccines and Preventive Services in Healthcare

Marta J. Ventura, Penn State University, State College, PA, United States

The first research problem is a new mathematical framework (integer linear program) that proposes to generate an optimal COVID-19 vaccination schedule for residents (in a region). This mathematical model coordinates all the necessary activities for selecting the locations and capacities of temporary medical facilities, purchasing single-dose and two-dose vaccines by medical facilities in an orderly manner, and generating a vaccination timetable for residents while minimizing the overall cost.

The second research problem is dental healthcare problem that estimates the impact of changes in Medicaid Dental Benefits on functional dentition in US Adults. We estimate the impact of the change in dental benefits due to Medicaid expansion in US states on tooth retention in 2012 and 2018 among working-aged adults who were Medicaid-eligible and not-eligible. We used a difference in difference model to examine the impact of changing dental benefits from a state's Medicaid policy on tooth retention.

The third research problem is about latent Gaussian Mixture Modeling (GMM) of Spatio-temporal COVID-19 Dynamics. We investigate the generation of dynamic COVID-19 density maps, modeling the multimodal probability density function (PDF) of a COVID-19 density maps, and tracking the epicenter of COVID-19 peaks. We present a novel approach to simulate the COVID-19 cases for statistical estimation and modeling of the multimodal PDF of a COVID density map.

The fourth research problem is a vehicle routing problem (VRP) application for an effective large-scale vaccine distribution for the state of Pennsylvania. We used state-of-the-art 'VRPsolvereasy' which is a Branch-Cut-and-Price exact solver for solving VRPs.

3 - Revealing Gaps in Modeled Smoking Histories: Implications for Lung Cancer Screening Eligibility

Hesam Mahmoudi, MGH Institute for Technology Assessment, Harvard Medical School, Boston, MA, United States, Alexandra L. Potter, Jacopo Di Silverstre, Andrew Eckel, Vidit Munshi, Chi-Fu Jeffrey Yang, G. Scott Gazelle

Purpose: Evaluate the National Cancer Institute's Smoking History Generator (SHG) for modeling smoking histories and assess its impact on lung cancer screening eligibility, with particular attention to how smoking duration may be undervalued compared to pack-year thresholds.

Methods: We analyzed data from the Southern Community Cohort Study (SCCS) and the Multiethnic Cohort (MEC) Study—both with substantial representation of minority groups—comparing SHG-generated smoking histories with real-world patterns. Discrepancies were assessed using smoking duration, pack-years, and proportions meeting established and proposed screening guidelines.

Results: Notable differences emerged between SHG simulations and observed data for both Black and White subpopulations. A higher proportion of Black smokers reported fewer than 20 pack-years in actual data than the SHG predicted, while a smaller proportion of White smokers reported fewer than 20 pack-years than indicated by SHG. As a result, reliance on SHG overestimates screening benefits for Black smokers under pack-year-based eligibility criteria and underestimates them for White smokers. Additionally, SHG consistently underestimated eligibility based on smoking duration thresholds, suggesting that exclusive focus on SHG simulation results can overshadow the potential advantages of duration-based criteria for all subpopulations.

Conclusions: These findings indicate that unstratified, pack-year-based SHG simulations may not accurately capture real-world smoking histories. Incorporating more detailed, stratified data into SHG—along with giving greater consideration to smoking duration—could enhance the accuracy of screening models and lead to more effective and broadly applicable screening policies.

4 - Improved Operations of Infectious Disease Elimination: a Model-Based Study of Polio

Yuming Sun, Georgia Institute of Technology, Atlanta, GA, United States, Pinar Keskinocak, Stephanie Kovacs, Lauren Steimle

Despite the availability of a new vaccine, persistent poliovirus outbreaks remain a global health concern, revealing the need to improve outbreak response vaccinations. Outbreak response aims to quickly vaccinate all at-risk children when outbreaks occur, regardless of their vaccination history. However, with limited vaccine supply, more strategic use of vaccines is essential, e.g., prioritizing under-vaccinated children. This study investigates whether operational improvements of outbreak response, specifically the vaccine allocation strategies, can achieve the elimination of poliovirus.

Developing a flexible deterministic compartmental model, we simulated outbreaks in Nigeria and evaluated several response scenarios. These varied in the number of vaccination rounds and the vaccine allocation strategies: proportional distribution, prioritization of under-vaccinated children, and prioritization of low-immunity children. Results showed that elimination was only achieved in the most sophisticated and idealized scenario incorporating five vaccination rounds all targeting truly low-immunity children. More practical scenarios, such as prioritizing the under-vaccinated, reduced transmission but did not eliminate the virus.

These findings highlight a key insight: although operational improvements in outbreak response can potentially eliminate poliovirus, the required changes may exceed the capacity of the global polio eradication program. To achieve and sustain elimination, a shift is needed from purely reactive outbreak response to strategies that incorporate

preventive vaccinations, such as adding the new vaccine to routine immunization. Our study contributes to global polio eradication efforts by identifying inefficiencies in outbreak response and emphasizing the need for prevention-oriented vaccination planning.

SC12

Building A Level 3 A314

Transportation on Networks

Job Market Showcase

Job Market Showcase

Chair: Daniela Cantarino, University of South Florida, Tampa, FL, United States

1 - Adapting Strategies for Trucking Platforms in the Gig Economy

Himani Ananthula, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Milind Sohoni, Achal Bassamboo

We examine how trip duration, shaped by network structure, affects labor supply in gig-based long-haul trucking. Using data from a large trucking platform in India, we show that longer trips reduce driver reliability. Further, using a queuing-based analytical model, we find that under demand uncertainty, adaptive dynamic staffing makes the Relay Network (RN) more cost-efficient than the Point-to-Point (P2P) model. Our results highlight how the dynamics of the on-demand economy, combined with supply behavior, shift optimal network design towards a win-win outcome.

2 - Toward a Hyperconnected Multi-Carrier Multi-Hub LTL Freight Transportation System

Sixtine GUERIN, Georgia Institute of Technology, Atlanta, GA, United States

3 - Cross- vs. in-Region Courier Routing in on-Demand Delivery

Arseniy Gorbushin, Rotman School of Management, Toronto, ON, Canada, Ming Hu, Xiaotang Yang, Yun Zhou

On-demand delivery platforms have expanded rapidly. A key advantage often attributed to their success is the ability to optimize courier routing by sharing couriers across locations and among multiple vendors (e.g., restaurants and grocery stores). However, in practice, couriers on food delivery platforms tend to work in the same regions, thus limiting the potential benefits of cross-region operations. To examine whether and how sharing couriers across regions benefits a firm, we develop a spatial queuing model where couriers serve as servers, and the delivery time depends on a customer's location, their chosen vendor, and the dispatch policy. Surprisingly, we find that in many cases, the in-region policy—under which couriers are dedicated to specific vendors—can outperform the cross-region policy, which uses a shared fleet to serve multiple vendors. This result stems from the randomness and potential imbalance in courier allocation that cross-region routing introduces. With an exogenous delivery fee, commonly used in the early stages of market penetration, the cross-region policy is more profitable when the market size is large; otherwise, the in-region policy is superior. However, with an endogenous delivery fee—typical for well-established firms who adjust the delivery fees to influence demand—the in-region policy is even more likely to be optimal: In addition to market size condition, high enough service value is required for cross-region policy to be efficient. Furthermore, we show that the cross-region policy perform better in markets where restaurants offer distinctive cuisines, whereas the in-region policy is more advantageous when restaurants have similar offerings.

4 - Optimizing Ambulance Deployment: a Stochastic-Machine Learning Framework for Emergency Service Efficiency

Daniela Cantarino, University of South Florida, Tampa, FL, United States, Jorge Acuna, Jose Zayas-Castro

Emergency department (ED) overcrowding remains a persistent challenge in the U.S. healthcare system, contributing to higher mortality rates, financial losses, and ambulance diversions. To mitigate these issues and enhance patient outcomes, hospital workflows, and system efficiency, centralized ambulance deployment strategies are critical to managing this flow effectively. This study proposes an integrated framework that combines multistage stochastic programming with machine learning to optimize ambulance allocation from accident sites to hospitals. We employ the progressive hedging algorithm to solve the allocation problem across diverse scenarios, incorporating stochastic demand variations, dynamic hospital occupancy, resource capabilities, wait times, travel distances, and patient acuity levels. The algorithm generates a robust dataset of optimal dispatch decisions, accounting for uncertainty in real-world conditions. Using this data, we train a machine learning model to predict near-optimal allocation strategies rapidly, enabling real-time decision-making. Preliminary results using real-world data from a county's Emergency Medical Services (EMS) demonstrate that our approach enhances resource utilization, reduces ambulance turnaround time, patient wait times, and improves system resilience compared to traditional static deployment methods. By bridging stochastic optimization and predictive modeling, this framework offers a scalable, adaptive solution for ambulance management, with potential to alleviate ED overcrowding and inform healthcare policy.

SC13

Building A Level 3 A315

Active Learning and Bandit Algorithms

Job Market Showcase

Job Market Showcase

Chair: Alexander Terenin, Cornell University, Ithaca, NY, United States

2 - What Does Thompson Sampling Optimize?

Yanlin Qu, Columbia University, New York, NY, United States, Hongseok Namkoong, Assaf Zeevi

Thompson Sampling and its variants are one of the most widely used algorithms for multi-armed bandit problems, known for their conceptual simplicity and strong empirical performance. Despite its popularity, a fundamental question about this heuristic algorithm remains elusive: What objective, exactly, is Thompson Sampling optimizing? We address this question by "rediscovering" Thompson Sampling as an online optimization algorithm that, at each round, minimizes a suitable notion of instantaneous regret regularized adaptively by a simple uncertainty measure. Our optimization perspective provides a fresh principled way to understand Thompson Sampling, and leads to natural avenues for algorithmic improvement.

3 - An Active Learning Framework for Multi-Group Mean Estimation

Abdellah Aznag, Columbia University, New York, NY, United States, Rachel Cummings, Adam Elmachtoub

We study a fundamental learning problem over multiple groups with unknown data distributions, where an analyst would like to learn the mean of each group. Moreover, we want to ensure that this data is collected in a relatively fair manner such that the noise of the estimate of each group is reasonable. In particular, we focus on settings where data are collected dynamically, which is important in adaptive experimentation for online platforms or adaptive clinical trials for healthcare. In our model, we employ an active learning framework to sequentially collect samples with bandit feedback, observing a sample in each period from the chosen group. After observing a sample, the analyst updates their estimate of the mean and variance of that group and chooses the next group accordingly. The analyst's objective is to dynamically collect samples to minimize the collective noise of the estimators, measured by the norm of the vector of variances of the mean estimators.

We propose an algorithm, Variance-UCB, that sequentially selects groups according to an upper confidence bound on the variance estimate. We provide a general theoretical framework for providing efficient bounds on learning from any underlying distribution where the variances can be estimated. This framework yields upper bounds on regret that improve significantly upon all existing bounds, as well as a collection of new results for different objectives and distributions than those previously studied.

4 - Bayesian Algorithms for Adversarial Decision-Making Under Uncertainty in General Action Spaces

Alexander Terenin, Cornell University, Ithaca, NY, United States

Understanding how to make decisions under uncertainty in a manner that guarantees strong performance in hindsight - no matter what might happen in the external environment - is a fundamental research question. The online learning game is a minimax game which captures this question, and is known by many names, including no-regret learning, and prediction with expert advice. Algorithms which achieve a no-regret online learning guarantee are a basic building block of other decision-making and equilibrium-computation algorithms, including certain forms of sample-efficient reinforcement learning. Most approaches are based on mirror descent variants such as exponential weights, which are convex-analytic in nature and best understood for discrete action spaces.

Given that online learning is a minimax game, it appears to have little to do with Bayesian decision-making. In spite of this, we show that a Bayesian approach - specifically, Thomson sampling over the space of the adversary's future actions (as opposed to the more-obvious state space) - admits a no-regret online learning guarantee, even in non-discrete action spaces. This is achieved by developing a Bayesian way of analyzing minimax regret of Thompson sampling, by decomposing regret into the regret the learner expects to incur according to the prior, and excess regret resulting from the presence of an adversary. With this, we show - in both the classical discrete setting, and on a unit cube against a bounded Lipschitz adversary - that Thompson sampling achieves a low excess regret, and therefore a low minimax regret. We conclude by discussing progress toward extending this to more general smoothness classes.

SC14

Building A Level 3 A316

Topics in Assortment Optimization

Job Market Showcase

Job Market Showcase

Chair: Shuo Sun, University of California, Berkeley, Berkeley, CA, United States

1 - Multi-Objective Assortment Optimization: Profit, Risk, Customer Utility, and Beyond

Zhen Chen, Arizona State University, Tempe, AZ, United States, Heng Zhang, Hongmin Li, Scott Webster

Assortment optimization is a fundamental problem in revenue management, in which the objective usually is to selectively offer a subset of products to maximize expected revenue or profit. However, business practices often involve multiple and potentially conflicting goals. In this work, we propose a general framework and a novel reformulation method for solving multi-objective assortment optimization problems. Specifically, we consider assortment problems whose objectives are sums of multiple convex objective functions on linear combinations of choice probabilities, and we present a reformulation that effectively “linearizes” the problem. We prove that the reformulated problem is equivalent to the original problem and that it leads to a unified solution approach to multi-objective assortment optimizations in various contexts. We show that the approach encompasses a wide range of operational objectives, such as risk, customer utility, (transformations of) market share, costs with economies of scale, and dualized convex constraints. We first illustrate our approach with the multinomial logit model, and then show that our framework leads to tractable solutions under the Markov chain choice model. Using large-scale numerical experiments, we highlight that our work provides a powerful and flexible tool for solving multi-objective assortment problems, which arise frequently in practice.

2 - Supply Chain-Driven Assortment Planning Under Uncertainty

Jisoo Park, Georgia Institute of Technology, Atlanta, GA, United States

We propose an integrated optimization framework for assortment planning that incorporates supply chain considerations using a rank-based customer choice model. Our multi-stage stochastic model captures multiple sources of uncertainty and customer substitution behavior, maximizing expected profit while aligning product offerings with operational capabilities. Empirical validation using data from an e-commerce furniture retailer demonstrates that our proposed approach significantly improves profitability and enhances supply chain resilience.

3 - Leveraging Assortment Similarities for Data-Driven Demand Prediction

Amin Farzaneh, University of Texas at Dallas, Richardson, TX, United States, Tarek Abdallah, Ashwin Venkataraman

Choice models are fundamental tools in revenue management, used to predict customer demand and optimize operational decisions such as pricing and assortment. Parametric models, like the MNL and nested logit, facilitate efficient estimation and decision-making but are prone to misspecification, which can degrade prediction accuracy. Nonparametric models, such as rank-based models, better capture complex substitution patterns and offer more accurate demand predictions, but are computationally intensive, harder to optimize, and require large datasets to avoid overfitting. This trade-off complicates the selection of an appropriate choice model in practice. In this work, we propose a fully data-driven framework that combines the strengths of both parametric and nonparametric approaches. Given access to historical transaction data, our method predicts choice probabilities for a test assortment by leveraging data from historical assortments that are most “similar” to it, drawing inspiration from nearest-neighbor methods in machine learning. We show that our method provides universal consistency of both demand predictions and assortment decisions—that is, they converge to the true values as historical data grows, regardless of the underlying choice model. Numerical experiments on synthetic data demonstrate the superior prediction and decision performance of our framework compared to parametric models like MNL and LC-MNL. We further introduce a hybrid approach that combines predictions from our similarity-based method with those from parametric models, enhancing prediction accuracy even with limited transaction data. This integration bridges the strengths of both approaches, making our framework practical and robust in real-world applications.

4 - A Markovian Approach for Cross-Category Complementarity in Choice Modeling

Shuo Sun, UC Berkeley, Berkeley, CA, United States, Rajan Udawani, Omar El Housni

In modern retail and e-commerce, consumers frequently purchase multiple items across different product categories, exhibiting both substitution and complementary behaviors. We propose a general framework for complementarity in settings where customers first select from a primary product category followed by secondary product(s), and the transition probabilities between product categories reflect complementarity. In contrast to previous approaches, our framework supports a tractable algorithm for joint assortment optimization across product categories. We demonstrate the efficacy of our framework by conducting extensive numerical experiments on both synthetic data and a large-scale transaction-level grocery dataset. We also introduce an empirical metric to quantify complementarity between different categories. Results show that the proposed framework improves model fit, prediction accuracy, and expected revenue when cross-category complementarity is present.

SC15

Building A Level 4 A401

ATS Best Paper and Best Dissertation Award

Award Session

Air Transportation Section

Chair: Felipe Delgado, Pontificia U Catolica- Chile, Santiago, Chile

Co-Chair: Zhenyu Gao, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

1 - Data-Driven Optimization for Air Traffic Flow Management with Trajectory Preferences

Guglielmo Lulli, Lancaster University, Lancaster, United Kingdom

In this talk, I present a novel data-driven optimization approach for trajectory-based air traffic flow management (ATFM). A key aspect of the proposed approach is the inclusion of airspace users' trajectory preferences, which are computed from traffic data by combining clustering and classification techniques. Machine learning is also used to extract consistent trajectory options, whereas optimization is applied to resolve demand-capacity imbalances by means of a mathematical programming model that judiciously assigns a feasible four-dimensional trajectory and a possible ground delay to each flight. The methodology has been tested on instances extracted from the Eurocontrol data repository. With more than 32,000 flights considered, we solve the largest instances of the ATFM problem available in the literature in short computational times that are reasonable from the practical point of view. As a by-product, we highlight the trade-off between preferences and delays as well as the potential benefits. Indeed, computing efficient solutions to the problem facilitates a consensus between the network manager and airspace users. In view of the level of accuracy of the solutions and the excellent computational performance, we are optimistic that the proposed approach can make a significant contribution to the development of the next generation of air traffic flow management tools.

2 - An Optimization Approach for the Terminal Airspace Scheduling Problem

Nuno Ribeiro, Singapore University of Technology and Design, Singapore, Singapore, Wayne Ng, Diana Jorge

Efficient management of the Terminal Manoeuvring Area (TMA) is vital to reduce delays, fuel burn, and emissions. While most research has centered on runway sequencing, the broader Terminal Airspace Scheduling Problem (TASP) remains less explored. We introduce **TMAOpt**, a metaheuristic that jointly optimizes runway sequencing and TMA decisions such as runway assignment, speed control, holding patterns, vectoring, and point merges. By combining Linear Programming with metaheuristics, TMAOpt delivers rapid feasible solutions within seconds and converges within minutes. Using real-world data from Changi Airport, results show delays reduced by up to 27% compared with state-of-the-art microsimulation tools. Sensitivity analyses further highlight the value of balanced TMA interventions.

3 - Advancing Urban Air Mobility: Strategic, Tactical, and Human-Centered Approaches to Traffic Management

Shulu Chen, George Washington University, Washington, DC, United States

Urban air mobility (UAM) offers a transformative solution for urban transportation. However, its integration faces unprecedented regulatory and operational challenges. Existing airspace frameworks are ill-equipped for high-density urban operations, requiring stringent safety measures to manage eVTOL aircraft over populated areas. Addressing these challenges demands a focus on both regulatory adaptation and operational efficiency.

This dissertation proposes an integrated traffic management framework for UAM that balances safety and efficiency by increasing autonomy in both vertiport and airspace operations. At the strategic level, a MILP based demand-capacity balancing (DCB) algorithm computes optimal departure schedules using ground delay to manage traffic bottlenecks. At the tactical level, a multi-agent reinforcement learning (MARL) approach provides real-time speed advisories to maintain safe separations. To further enhance performance across varying scenarios, a LLM-augmented MARL framework adjusts reward structures based on contextual needs and human preferences, improving conflict resolution efficiency. To bridge research with practical deployment, this work introduces the vertiport human-automation teaming toolbox (V-HATT), a decision-support system that integrates human operators with automation.

By combining strategic and tactical conflict management with human-autonomy teaming, this research delivers a scalable, adaptive, and safety-focused framework for UAM operations. These contributions lay a foundation for the sustainable and efficient growth of future urban air transportation systems.

SC16

Building A Level 4 A402

Optimal Sourcing at Amazon

Invited Session

The Practice Section of Informis

Chair: Ozge Sahin, Johns Hopkins Carey Business School, Baltimore, MD, United States

1 - Just-in-TIME-Aware Long Lead Buying Policy

Yuguang Wu, Amazon, Bellevue, WA, United States, Muhong Zhang, Xiuli Chao, Huseyin Topaloglu

Amazon Retail employs a dual-sourcing strategy for many products, leveraging short-lead time Just-In-Time (JIT) and long-lead time Direct Import (DI) channels to balance inventory replenishment cost and risk. However, managing dual-sourcing inventory replenishment poses significant challenges due to lost-sales dynamics, non-stationary and correlated demand, and stochastic supply availability. Existing analytical methods offer theoretical rigor but face practical limitations, including computational inefficiency and restrictive assumptions such as i.i.d demand or demand backlogging. To address these challenges, we propose the JIT-Aware Long-Lead Buying (JITA-LLB) algorithm, a simulation-based optimization framework that integrates long-lead buying decisions with JIT policies. The algorithm employs sample-path gradients to iteratively optimize multi-period replenishment strategies, offering scalability and adaptability to real-world complexities. We evaluate JITA-LLB against heuristic TIP-scaling policies using real-world test instances, demonstrating its competitive expected long term profit performance and its ability to adapt to high demand variability. This work advances scalable dual-sourcing inventory management through a practical and modular approach.

2 - Getting Abstract to Me Week 9/15 -Ailey

Thiago Mosquero, Amazon, Bellevue, WA, United States

Getting abstract to me week 9/15 -Ailey

3 - Multi-Period Optimization Models for Distribution Center Inventory Management Under Uncertainty

Caner Taskin, Amazon, Seattle, WA, United States, Alp Muharremoglu

This research presents optimization-based approaches for managing buying decisions in distribution centers (DCs) within Amazon's e-commerce network. Our models are designed to handle multi-period planning horizons while considering various operational constraints and uncertainties in demand and supply. Our models aim to improve inventory management efficiency while maintaining high service levels. We explore different mathematical modeling approaches to address the inherent complexities of DC inventory management and underlying uncertainties while maintaining computational tractability considering Amazon's large-scale network.

4 - Structure-Informed Deep Reinforcement Learning for Inventory Management

Alvaro Maggiar, Amazon.com, Arlington, VA, United States

This paper investigates the application of Deep Reinforcement Learning (DRL) to classical inventory management problems, with a focus on practical implementation considerations. We apply a DRL algorithm based on DirectBackprop to several fundamental inventory management scenarios including multi-period systems with lost sales (with and without lead times), perishable inventory management, dual sourcing, and joint inventory procurement and removal. The DRL approach learns policies across products using only historical information that would be available in practice, avoiding unrealistic assumptions about demand distributions or access to distribution parameters. We demonstrate that our generic DRL implementation performs competitively against or outperforms established benchmarks and heuristics across these diverse settings, while requiring minimal parameter tuning. Through examination of the learned policies, we show that the DRL approach naturally captures many known structural properties of optimal policies derived from traditional operations research methods. To further improve policy performance and interpretability, we propose a Structure-Informed Policy Network technique that explicitly incorporates analytically derived characteristics of optimal policies into the learning process. This approach can help interpretability and add robustness to the policy in out-of-sample performance, as we demonstrate in an example with realistic demand data. Our work bridges the gap between data-driven learning and analytical insights in inventory management while maintaining practical applicability.

5 - Pareto Frontier Profiling for Order Assignment Optimization at Amazon

Neo Huang, Amazon, San Diego, CA, United States, Sherief Reda, Jikai Zou, Andrea Qualizza

We will discuss an application of the Augmented ϵ -Constraint Method (AUGMECON) to profile the Pareto Frontier and generate multi-objective weights (MOW) for the order assignment problem at Amazon responsible to decide how and when each order gets fulfilled. This approach uses stratified sampling and symmetry reduction to

consolidate similar demand attributes, significantly reducing the problem complexity. Unlike other approaches, our framework generates MOWs for production use directly from Pareto-efficient solutions without requiring reference solutions and enables weight generation at any desired frontier location within predefined thresholds.

SC17

Building A Level 4 A403

Inbound Supply Chain Planning and Execution

Invited Session

Practice Track

Chair: Daniel Ulch, Amazon, Bellevue, WA, United States

Co-Chair: Marcus Poggi, Amazon, Bellevue, WA, United States

1 - Amazon Distribution Centers Picking Optimization

PREM KUMAR VISWANATHAN, Amazon, Bellevue, WA, United States, Shuocheng Guo

This study introduced novel optimization-based algorithms for order picking in bulk businesses, focusing on Amazon Distribution Centers (DCs) that supply downstream warehouses and physical stores, with the goal of optimizing order fulfillment and transportation costs. While existing methods optimize static-allocated order picking with fixed demand-to-inventory allocation, they often yield suboptimal solutions by failing to address dynamically-allocated orders stored across multiple locations.

In this study, we propose a two-stage optimization-based algorithm that efficiently handles both static and dynamic-allocated orders, addressing current DC operations while maintaining flexibility and generalization for future requirements. Given that over 70% of DC orders are static-allocated, our algorithm strategically decomposes the problem into two stages: first, assigning items/SKUs with static locations to containers using an enhanced version of the current prototype, and second, allocating remaining items/SKUs with multiple inventory locations to partially-filled containers.

To improve computational efficiency, we introduce two acceleration techniques, Pre-fixing and Pre-assignment, which integrates detailed warehouse layout information to model realistic travel time distributions. Specifically, we account for both horizontal movement along aisles and vertical movement for accessing higher bins. This two-dimensional travel pattern, where pickers must walk horizontally and operate their order picker lift truck vertically, introduces complexity in computations. Extensive numerical experiments across 167 instances demonstrate about 40% average improvement in run times over more optimal formulations for medium and larger instances.

2 - Inbound Inventory Allocation in a Tiered Fulfillment Network with Storage Distribution Centers

Di Wu, Amazon.com, Seattle, WA, United States, Tolga Cezik, STEPHEN GRAVES, Daniel Chen, Jason Acimovic

With the scale of Amazon's fulfillment network, optimally placing incoming inventory into fulfillment nodes becomes challenging due to the large decision space, numerous physical capacity constraints, and operational challenges. Existing processes handling inbound decisions, although sophisticated, are also complex. As a result, managers cannot audit the processes without spending significant time 'deep-diving'. Because these processes lack auditable transparency, the inbound system operates sub-optimally in ways we cannot detect. Inspired by the recent success in regionalizing outbound, this paper proposes a simplified, holistic inventory allocation policy based on a tiered fulfillment network that partitions demand areas into regions. In this regime, Amazon must distribute inventory at each echelon level to the lower levels in its region, one after the other, while at the last level, Amazon must split inventory between customer-facing fulfillment centers and pure storage nodes. We focus on developing inventory policies that perform well in such a tiered network, from both cost and computational efficiency points of view.

3 - When to Split: Optimizing Container Breaking and Inventory Distribution Decisions in an Inbound Supply Chain

Marcus Poggi, Amazon, Bellevue, WA, United States, Daniel Ulch

This talk presents a novel approach to solving the Placement Adherence (PLAD) optimization problem, which minimizes deviation from inventory targets at warehouse nodes while respecting resource constraints in complex supply chain networks. We focus on the computational challenges of this problem and introduce an efficient column generation heuristic that significantly improves scalability. The Placement Adherence problem is particularly challenging due to its size and complexity - involving millions of unique products flowing through a multi-echelon network with various resource constraints. Products travel through the network in batches (e.g. pallets, cases, totes). Breaking batches down enables a higher degree of inventory spread, but consumes more resources to sort at downstream nodes. Resource constraints link the products together, creating a massive integer multi-commodity generalized flow problem. Our column generation approach reformulates the problem to select, for each product, a feasible integer flow through the network. The master problem coordinates resource allocation across products. This decomposition approach provides several advantages: it yields stronger bounds than direct linear relaxation, scales to hundreds of thousands of products (compared to just thousands with traditional MIP solvers), and produces high-quality solutions that respect the integer nature of case breaking decisions. We'll discuss implementation challenges, including parallelization strategies and efficient subproblem solution techniques.

SC18

Building A Level 4 A404

Behavioral Modeling and Learning in Smart Mobility Systems

Invited Session

TSL: Intelligent Transportation Systems

Chair: Devansh Jalota, Columbia University, Severna Park, MD, United States

1 - Analyzing Platform-Worker Dynamics in the Gig Economy: A Simulation-Based Approach

Tara Eslami, Northeastern University, Boston, MA, United States, Ozlem Ergun

The gig economy has changed the traditional working by enabling short-term, flexible employment. Although workers have autonomy, algorithmic control and changing market conditions bring them various challenges. Changes in policies such as pricing and task allocation can affect workers' behavior and well-being. However, most of the research is aimed at improving the system efficiency or customer satisfaction, with little consideration for how the workers are affected. Even where workers' behavior is being simulated, it is presented as a constraint, not an outcome of value. This leaves a gap for understanding the complete dynamics of gig economy.

These changes within the system and policies of the gig economy impact both the workers welfare and the platform performance. We seek to create a more comprehensive picture of these markets with efficiency as well as fairness. Using an agent-based model (ABM), we simulate heterogeneous workers and an adaptive platform within a dynamic environment. The workers have varied strategies - maximizing earnings, following their schedules, or targeting a specific income. They also adapt their behavior based on the market conditions. The platform assigns jobs, determines wages, and adjusts policies depending on system performance.

Our simulations demonstrate how policy interventions affect the workers and changes the system outcomes. We can see how some workers leave the system early and decrease peak-hour reliability, while others provide stability at the cost of responsiveness. Through this model, platform designers, policymakers, and scholars can better explore how sustainable, efficient, and fair gig systems can be created for all stakeholders.

2 - Meal Delivery Routing with Courier Behavior and Graph Neural Networks

Niels de Jong, University of Groningen, Groningen, Netherlands, Ilke Bakir, Kees Jan Roodbergen

We address the Meal Delivery Routing Problem (MDRP) faced by online food delivery platforms, where customer orders arrive dynamically and must be assigned to autonomous couriers. Each courier picks up orders from merchants and delivers them to customers. The platform must handle several challenges: couriers have varying behaviors and capabilities; courier-to-order assignment decisions must balance immediate and future costs; decisions must be made in seconds; multiple orders can be efficiently bundled; some assignments can be delayed; and there is inherent uncertainty in food preparation times, travel speeds, and order delivery locations. To address

these challenges, we develop a scalable, polynomial-time heuristic using Graph Neural Networks (GNNs) to estimate future assignment costs in a rolling horizon assignment framework. We test our approach using a modular, open-source digital twin that simulates millions of deliveries in seconds and is validated on real data from Meituan. Our GNN-based heuristic is trained using Twin Delayed Deep Deterministic Policy Gradient (TD3). Results show our method reduces average delivery time (click-to-door) by 11.6% and order rejections by 2% compared to a myopic benchmark. It also adapts to varying courier behaviors and changing spatio-temporal demand, enabling real-time, efficient decision-making for food delivery and related applications like ride-sharing.

3 - Understanding the Behavior of Large Language Models: a Case Study on the Value of Travel TIME

Yingnan Yan, University of Michigan, Ann Arbor, MI, United States, Tianming Liu, Yafeng Yin

In recent years, the rapid emergence of artificial intelligence (AI) has greatly accelerated the evolution of modern transportation systems. The latest AI paradigms exhibit unprecedented intelligence. Large language models (LLMs), the state-of-the-art AI models, excel at language processing, contextual interaction, and decision-making, behaving much like humans. AI models like LLMs are increasingly deployed in transportation systems as agents for human, which may deeply engage with humans and be applied to analyses of human behavior. Therefore, it is essential to assess how closely the behaviors of LLMs align with humans in transportation scenarios. Currently, the investigation into LLMs' behaviors in transportation still remains limited and vague. This study offers a quantitative comparison of behavioral patterns between LLMs and humans through the estimation of LLMs' value of travel time (VOT). Results show that LLMs can differentiate the assigned context and display human-like responses correspondingly, though the magnitudes of these distinctions are smaller than those of humans. Overall, the behaviors of LLMs do resemble human travelers, suggesting considerable potential for having LLMs coexist with humans in transportation systems and using LLMs in pilot studies of human behavior.

Keywords: Large language model, Behavior in transportation, Human agent

4 - Distributional End-to-End Learning of User Equilibrium with Uncertainty Quantification

Peisen Li, University of Michigan, Ann Arbor, MI, United States, Zhichen Liu, Yafeng Yin

Traditional user equilibrium models often rely on oversimplified assumptions about traveler behavior and produce point predictions without reflecting real-world uncertainties. My research addresses these limitations by extending the end-to-end learning framework to incorporate distributional prediction and uncertainty estimation. This approach integrates Monte Carlo dropout and recursive neural network refinement to produce a distribution over UE flow predictions. Instead of deterministic outputs, the model estimates both the mean and variance of predicted traffic flows, offering richer information about confidence levels under varying input conditions (e.g., demand fluctuations, network states). A recursive network architecture refines UE predictions iteratively, while dropout layers in the Attribute Net and Weight Net enable stochastic sampling. This combination allows for the estimation of both model and data uncertainty. The proposed framework is evaluated using the Sioux Falls transportation network, where performance is measured using MAPE and PICP metrics. Results demonstrate that our method not only improves predictive accuracy but also yields well-calibrated uncertainty estimates, which are critical for robust transportation planning and policy evaluation.

5 - Combatting Fraud in Welfare Programs: Theory and Applications in Smart Mobility

Devansh Jalota, Stanford University, Stanford, CA, United States, Matthew Tsao, Marco Pavone

As AI-enabled algorithmic decision-making systems become more prevalent, there is a growing need for market mechanisms resilient to user fraud and strategic behavior. Of particular interest in this work is the concern of misreporting fraud in welfare programs, wherein users can misreport their private attributes to receive more aid than they are entitled to, often resulting in millions of dollars of financial losses and reduced aid for eligible low-income users. To combat this concern, in this work we develop an audit mechanism to mitigate misreporting fraud and present algorithms for societal-scale equilibrium computation in the signaling game it induces, with a real-world use case in law enforcement in a smart mobility setting. In our proposed audit mechanism, the administrator can audit users at some cost and levy fines against them for misreporting. We study the resulting signaling game induced by our audit mechanism in two settings: a static setting where the distribution of user types is common knowledge, and an online setting where the distribution of user types is not known to the administrator. In the static setting, we present a linear programming approach to compute the resulting equilibria and, in the online setting, we develop an auditing algorithm that achieves sublinear regret with respect to the optimal static strategy. Finally, we highlight the practical viability of our audit mechanism in mitigating misreporting fraud through a case study based on Washington D.C.'s federal transit benefits program.

SC20

Building A Level 4 A406

Network Analysis for Humanitarian Logistics

Invited Session

Telecommunications and Network Analytics

Chair: Abdullah Konak, Penn State Berks, Reading, PA, United States

1 - Improving Food Accessibility in Food Bank Supply Chains: Expanding Network Capacity Through Partner Agency Recruitment

Lauren Davis, North Carolina A&T State University, Greensboro, NC, United States, Rabiatsu Bonku, Funda Samanlioglu, Shubhra Paul

Food banks are non-profit organizations that distribute purchased and donated food to people in need of food assistance. Their distribution network consists of local hunger relief agencies (e.g. soup kitchens and food pantries) that work collaboratively with the food bank to ensure efficient, effective and equitable flow of food from food bank warehouses to food-insecure individuals. Partner agencies play a critical role in this supply chain and their physical location throughout the food bank's service area can contribute to inequitable distribution and low accessibility for certain populations. In this work, we aim to improve accessibility and equitable distribution by expanding network capacity through partner agency recruitment. Utilizing data from a local food bank, we identify geographic areas of unmet need. We also develop a framework for identifying potential agencies to recruit and utilize k-means clustering to estimate their service capacity. Lastly, we develop a capacitated set covering model to determine the optimal number of new agencies to on-board, taking into consideration food bank goals and preferences related to the target population and type of programs offered.

2 - Network Fairness Analysis in Food Distribution Systems

Sadan Kulturel-Konak, Pennsylvania State University, Berks, Reading, PA, United States, Abdullah Konak, Lauren Davis, Steven Jiang, Ryan Hassler

This study presents a practical framework designed to enhance fairness in the distribution of emergency food across the United States. It focuses on ensuring that vulnerable communities, particularly those with limited access to transportation, receive equitable support. The framework helps identify gaps and inefficiencies in current food distribution systems by combining local data, mapping tools, and mathematical optimization. A key feature of this approach is its strong collaboration with stakeholders, including food banks, community groups, volunteers, and the individuals receiving assistance. To enhance understanding and actionability of the data, the framework includes customized visuals and decision-support tools that do not require technical expertise. It also establishes feedback loops, allowing these tools to evolve over time based on what works and what doesn't. Ultimately, the goal is to support communities in developing more equitable, efficient, and resilient food systems. Real-world examples from a regional food bank will illustrate how this framework can be applied in practice.

3 - Food Insecurity and Charitable Food Systems: Analyzing Donations at a Local Food Bank

Muhammad Habeeb Mohammed, North Carolina A&T State University, Greensboro, NC, United States, Florielvis Pernaleta, Steven Jiang, Lauren Davis, Abdullah Konak, Sadan Kulturel-Konak

Food insecurity is a condition in which households lack consistent access to adequate, nutritious food. Food insecurity remains a critical global challenge, even in developed nations such as the United States. Despite federal food assistance programs, millions of Americans continue to experience food insecurity, highlighting the vital role of charitable food systems like food banks in hunger relief. While food banks universally face challenges such as securing consistent donations, each operates within unique local contexts. This study analyzes donation data from a local food bank to identify patterns and trends in food contributions. By examining donor behavior and seasonal fluctuations, the findings can inform operational decisions such as optimizing truck capacity for donation pickups and ultimately enhance the food bank's efficiency and impact.

4 - Multi-Criteria Network Consolidation Problem in Charitable Food Distribution

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

The growing and evolving demand for food assistance, coupled with limited and fluctuating food donation resources, presents significant challenges for food banks. As traditional sources of donations diminish, food banks must periodically reassess and optimize their distribution networks despite constrained operational capacities. This study addresses the strategic planning dilemma faced by food banks, which seek to streamline delivery operations by reducing, consolidating, or relocating service sites. Such changes, however, risk reducing food accessibility for vulnerable populations, particularly those with limited transportation options. We propose a network reconfiguration and consolidation framework that integrates principles of social justice and operational efficiency. Key objectives include equitable access, fair distribution, respect for rights and dignity, and financial sustainability. This multifaceted approach aims to support food banks in making informed, balanced decisions that align logistical efficiency with community well-being.

5 - Evaluating Food Accessibility: Analyzing the Impact of Fresh Mobile Markets Using the Enhanced Two-Step Floating Catchment Area Method

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

Food insecurity persists as a pressing issue in low-income and low-access (LI/LA) communities, where physical and economic barriers limit access to nutritious food. This study evaluates the impact of Fresh Mobile Markets (FMMs) on food accessibility in Greensboro, NC, using the Enhanced Two-Step Floating Catchment Area (E2SFCA) method. Unlike traditional travel-time-based measures, the E2SFCA method accounts for both the supply of food assistance facilities—FMMs and food pantries—and the demand from surrounding low-income households, offering a more robust spatial analysis. Accessibility was assessed at the U.S. Census block group level, mitigating the modifiable areal unit problem (MAUP) and ensuring more precise estimates. Comparative analysis between travel time to the nearest facility and E2SFCA-derived accessibility scores reveals disparities, with travel-time measures often overestimating access in densely populated, high-need areas. Preliminary findings suggest that integrating FMMs into the food assistance network enhances spatial accessibility, but gaps remain in North and Southeast Guilford County due to scheduling and location constraints. This study underscores the value of advanced spatial models in optimizing food assistance strategies and highlights opportunities for improving the equitable distribution of mobile market services.

SC22

Building A Level 4 A408

Decision Making and Experimentation

Invited Session

Applied Probability Society

Chair: Sadegh Shirani, Stanford University, Stanford, CA, United States

1 - Asymptotically Efficient Distributed Experimentation

Ankur Mani, University of Minnesota - Twin Cities, Minneapolis, MN, United States, Ilan Lobel, Josh Reed

Sequential decision making by a large set of myopic agents has gained significant attention over the past decade. In such settings, even a little amount of experimentation from a few agents would benefit all others but obtaining such experimentation could be challenging for a central planner. The academic literature has focused on mechanisms for promoting experimentation through monetary incentives and persuasion through careful information disclosure. In this paper, we study a simple control that the central planner can use to coordinate experimentation. We consider a set of myopic agents that observe their own histories but not the histories of other agents. In a continuous-time stochastic multi-armed bandit model, the agents pick arms myopically and receive instantaneous rewards. Meanwhile, the central planner can observe the history of all agents. We consider a class of policies where the central planner is allowed to irrevocably remove arms. We show that an appropriately chosen policy within this class can generate the needed experimentation and match the regret bounds for a centralized problem thus mitigating the cost of decentralization. We also quantify the minimum number of agents that are needed for such a policy to be asymptotically optimal and the impact of the number of agents on the speed of learning.

2 - Incentive as Instruments: Identification and Estimation of Treatment Effects with Unobserved Confounders In Treatment Choice

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

Unobserved Confounders are one of the most difficult challenges in estimating treatment effect in observational study. In this work, we model the selection of treatment as utility maximization (or more broadly, as certain threshold policy). Under this condition, we provide sufficient conditions for treatment identification. When it is not identified, we provide sharp identification sets, as well as efficiency bound for convergence to this identified set.

3 - Causal Message Passing for Experiments with Unknown and General Network Interference

Sadegh Shirani, Stanford University, Stanford, CA, United States, Mohsen Bayati

Randomized experiments are the gold standard for assessing the impact of various decisions, including policy changes, medical treatments, or product enhancements. However, these experiments often face the challenge of network interference, where a decision affecting one individual impacts others, complicating the analysis. Our research introduces Causal Message-Passing, a framework inspired by techniques from physics and statistics to tackle this complexity. It allows for a more nuanced understanding of interconnected data, crucial for predicting outcomes in environments where individual responses are interdependent. This approach enhances the accuracy of decision-making in networked settings, offering insights for scientists and practitioners across diverse fields.

SC23

Building A Level 4 A409

Applications in Optimization

Invited Session

Minority Issues Forum

Chair: Alexandra Newman, Colorado School of Mines, Golden, CO, United States

1 - Multi-Objective Optimization for Equitable Electric Vehicle Charger Siting

Kevin Bamwiso, Colorado School of Mines, Golden, CO, United States, Estifanose Shiferaw

As the adoption of electric vehicles (EVs) accelerates across the United States, the demand for accessible and reliable public charging infrastructure continues to rise. However, the market-driven placement of charging stations often prioritizes high-income, high-traffic urban areas, leaving under-served communities with limited access to the infrastructure needed to support EV ownership. This gap exacerbates existing environmental and transportation inequalities, making clean transportation adoption less attainable in these underserved communities. To address this challenge, we develop a tri-objective mixed-integer programming model that simultaneously considers social equity, user demand, and infrastructure costs in EV charging infrastructure placement. We maximize equity and demand while minimizing infrastructure costs. Equity metrics include the consideration of low-income and minority populations, as well as households without off-street parking while demand metrics consider regions

with high EV ownership potential, high-income households, and total vehicles. Finally, cost components address electrical grid constraints and financial feasibility. Studies in Seattle and Pittsburgh provide insights into how cities can expand EV infrastructure in a way that is both financially sustainable and socially inclusive, supporting a just transition to clean transportation.

2 - Exploring Equality and Sustainability Trade-Offs of Energy Transition Outcomes in the United States in 2050

Destenie Nock, Carnegie Mellon University, Pittsburgh, PA, United States, Teagan Goforth, Maxwell Brown, Tapajyoti Ghosh, Patrick Lamers

With increasing focus on equitable and just energy transition, it is critical to understand the trade-offs of different decarbonization outcomes across economic, environmental, and social sustainability criteria. In this analysis, we use a multi-criteria decision analysis to quantify sustainability outcomes across 32 decarbonization outcomes in 2050 in the U.S. The economic sustainability criteria we use are system cost, national average retail rate, and electricity system employment. The environmental sustainability criteria we use are life cycle greenhouse gas emissions, life cycle water depletion, life cycle land transformation, and air pollution fatalities. The social sustainability (distributional impacts) criteria we use are retail rate equality across states, electricity employment equality across low-income households, and air pollution disparities across census tracts. We evaluate performance across these criteria under eleven different stakeholder preference scenarios. We find that decarbonization policies with indefinitely extended tax credits have the highest sustainability score under equal criteria weighting, with greater investments in renewable energy technologies, and result in better environmental, system cost, job, and air pollution disparities compared to mid-case scenarios, that only include current policies and CO2 reduction targets. We also see that our multi-criteria decision analysis identifies decarbonization outcomes that would not have been identified as optimal under a single objective, which highlights the importance of trade-off analyses to understand decarbonization outcomes more holistically.

3 - Degrees of Equity: How Households Adapt with Energy Efficiency Interventions

Kester Wade, Carnegie Mellon University, Pittsburgh, PA, United States, Destenie Nock, Xue Gao

Energy efficiency and conservation play a central role in addressing the challenges of energy system transitions in the US. Powering society's growing needs would directly benefit from reduced energy costs—requiring less energy for units of output or activities. These strategies can also improve affordability for households to meet their energy needs. Yet, despite their importance for climate goals and energy affordability, barriers often prevent low-income households from benefitting. Energy poverty involves excessive spending on energy bills, disconnections, and sacrificing essential services like heating and cooling to limit costs. There's a lack of understanding regarding how energy efficiency programs impact vulnerable groups.

This study considers ~125,000 households in a Florida county and examines the impacts of energy efficiency and conservation programs on household heating and cooling behaviors over a 10-year period. We focus on the outdoor temperatures at which households turn on heating or air conditioning systems to understand whether low-income households tolerate extreme indoor temperatures to reduce financial stress. Difference-in-difference comparisons help us estimate how these household behaviors change during and after voluntary energy efficiency program participation such as loans or rebates for new efficient appliances.

Energy efficiency technologies and conservation behaviors have massive potential to improve physical and financial well-being. This especially affects households that might engage in risky behaviors to mitigate their energy needs. We advance the scientific understanding of whether, why, and how different energy efficiency and conservation programs work in different households, informing over 900 similar programs offered by U.S. utilities.

4 - From Data to Decisions: Optimizing Pediatric Vaccine Tender Schedules Under Uncertainty

Nicholas Uhorchak, University of Arkansas, Fayetteville, AR, United States, Ruben Proano, Sandra Eksioglu, Burak Eksioglu, Asif Sharon

This study presents a data-driven framework for long-term pediatric vaccine tender scheduling, enhancing existing procurement practices to address future uncertainty in demand for vaccines and manufacturing capacity, and improve long term decision-making. While UNICEF and similar agencies regularly organize multi-year tenders, additional tenders are often initiated in response to unexpected supply shortfalls, disease outbreaks, or production issues—leading to inefficiencies such as stockouts, excess inventory, and reduced vaccine coverage due to logistical and storage constraints. We propose a stochastic optimization model to generate multi-year tender schedules under uncertain demand for vaccines and manufacturing capacity. We develop three models that focus on affordability (balanced), profitability (max profit), and public health (social benefit) impact. We evaluate the impacts of these models with regards to optimal sequencing and timing of tenders, expected tender costs and expected number of missed doses. We propose an L-shaped algorithm to solve these models. We use Pareto-optimal cuts and valid inequalities to accelerate the convergence of the algorithm and ensure practical solution times. We use real-world data from UNICEF to model uncertainty in vaccine demand, and synthetically generate manufacturing capacity scenarios. This approach is critical to capturing the variability that affects long-term tender scheduling in practice. Results indicate that while the balanced model yields marginally more missed doses compared to a purely social objective, it significantly reduces administrative complexity for global health agencies such as UNICEF, PAHO, and GAVI.

SC24

Building A Level 4 A410

Operations Research to Help Address the Mental Health Crisis on College Campuses

TutORial session

TutORials

Chair: David Alderson, Naval Postgraduate School, Monterey, CA, United States

1 - Operations Research to Help Address the Mental Health Crisis on College Campuses

Hrayer Aprahamian, Texas A&M University, College Station, TX, United States

In this TutORial, we address the growing mental health crisis affecting university campuses across the nation. Specifically, we provide a comprehensive overview of the challenges faced by Counseling and Psychological Services (CAPS) centers and present a range of data analytics and Operations Research tools aimed at mitigating these issues. By posing key unresolved questions, we offer a holistic analysis that spans the spectrum from individual patient outcomes to system-level operational performance,

highlighting the complex interdependencies between clinical effectiveness and operational efficiency. Through this lens, we present a structured and accessible framework that both summarizes and contextualizes the primary operational and clinical challenges in university mental health services. In doing so, we also showcase a variety of data-driven methodologies designed to address these challenges, offering insight into the strengths and limitations of current approaches.

SC25

Building A Level 4 A411

Grand Challenges in Public Sector Operations Research: Opportunities at the Intersection of AI and OR

Panel Session

Women in OR/MS (WORMS)

Co-Chair: Irem Sengul Orgut, Loyola University Chicago, Tuscaloosa, AL, United States

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

1 - Moderator Panelist

Chaitra Gopalappa, University of Massachusetts, Amherst, Amherst, MA, United States

The OR community has a strong tradition of tackling socially engaged and impactful application areas, many of which are characterized by high levels of complexity, uncertainty, and evolving constraints. As public sector challenges—from health equity and infrastructure to emergency response and resource allocation—grow in scale and intricacy, there is an urgent need to explore new computational paradigms. Artificial Intelligence (AI), with its ability to learn from large-scale data and model complex interactions, offers powerful tools but often lacks the robustness and transparency required for high-stakes decision-making. Conversely, OR methods are valued for their rigor and reliability in decision support but may struggle to scale or adapt in dynamic environments.

This panel will explore grand challenges in the public sector that require the complementary strengths of AI and OR. Panelists will discuss where the integration of AI and OR creates new possibilities that neither field could achieve alone, identify current gaps and methodological needs, and chart future research directions that can deepen the societal impact of the AI+OR interface.

2 - Moderator Panelist

Irem Sengul Orgut, Loyola University Chicago, Chicago, IL, United States

TBD

3 - Panelist

Hamsa Bastani, Wharton School, Philadelphia, PA, United States

4 - Panelist

Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States

5 - Panelist

Lavanya Marla, University of Illinois at Urbana-Champaign, Urbana, IL, United States

6 - Panelist

Shalome Hanisha Anand Tatapudi, University of South Florida, Tampa, FL, United States

SC26

Building A Level 4 A412

Location for Social Good

Invited Session

Location Analysis

Chair: Marziehsadat Rezaei, University of Louisville, Louisville, KY, United States

Co-Chair: Monica Gentili, University of Louisville, Louisville, KY, United States

1 - Location-Based Planning for Mixed-Fleet Transit: Balancing Equity and Efficiency in Uneven Urban Demand

Madelaine Martinez Ferguson, University of Tennessee, Knoxville, TN, United States, Mustafa Camur, Xueping Li

Many urbanized areas in the US exhibit spatially and temporally sparse transit demand, posing a resource allocation challenge for transit agencies (TAs) due to the ridership–coverage dilemma. Agencies often prioritize high-ridership corridors at the expense of broader coverage, limiting accessibility for transit-dependent populations. As emerging mobility options such as shared autonomous vehicles (SAVs) gain traction, their integration into public transportation systems offers an opportunity to enhance both access and efficiency. However, research on mixed-fleet operations—combining buses and SAVs—remains limited. This study contributes to closing this gap in two significant ways. First, it involves a location decision to select underserved areas or routes for system implementation, guided by a transit coverage analysis. Second, it identifies transition points—bus stops—where each fleet type will operate based on spatial demand density. Buses are assigned to dense corridors, while SAVs are deployed in low-density areas. These decisions are evaluated through an agent-based simulation framework that captures operational dynamics. The model explores trade-offs between vehicle capacity and flexibility, offering insights into how integrated, adaptive transit configurations can enhance equity and service efficiency in urban settings with heterogeneous demand.

2 - Location and Mode Selection in Drone-Enabled Medical Logistics for Transplant Support

Kevin Martin, University of Missouri, St Louis, Wildwood, MO, United States, Shakiba Enayati, Haitao Li

Timely testing of donor blood samples is a critical component of the organ donation and transplantation process. Laboratory results directly inform organ allocation decisions that can save lives. In regions where transplant coordination spans large geographic areas, delays in transporting blood samples to central laboratories can jeopardize organ viability and matching efficiency. This challenge is especially acute in rural and underserved areas where traditional ground and air transportation can be slow and expensive. Emerging drone technologies offer a promising solution by enabling swift, reliable, and cost-effective delivery of time-sensitive medical specimens across long distances.

Our study proposes a mixed-integer linear programming (MILP) model to design an optimized multi-modal distribution network for donor blood sample transportation. The model determines the most efficient combination of transportation modes—including drones—along the delivery routes from hospitals to the laboratories, while selecting strategic locations for long-term drone infrastructure. The objective is to minimize transportation costs while satisfying constraints on maximum allowable transit time.

The model is designed in collaboration with an organ procurement organization located in Missouri. The geographic focus includes hospitals in the southwestern parts of the state and a testing lab in the St. Louis area. Current logistics rely on costly ground couriers and aircraft. Our model evaluates the integration of drones to improve efficiency, reduce transit times, and lower costs while complementing existing services. We also conduct sensitivity analyses and trade-offs for practical decision-making. This work contributes to accelerating the donor matching process and enhancing the overall effectiveness of organ transplant logistics.

3 - Location Optimization for Mobile Ground-Based Polarimetric Radar for Strategic Monitoring of Severe Storm Systems

Xin (Selena) Feng, University of Oklahoma, Norman, OK, United States, Bikram Parajuli, Chenchun Wang

Accurate radar observations are critical to improving the prediction and warning of life-threatening and economically impactful severe weather events. Compared to fixed ground-based radar systems, mobile radars have proven to offer better spatial and temporal resolution when they are deployed close to where the perception system occurs. A key issue about mobile radar observations that has been overlooked is where to deploy the mobile radar vehicles. The scientists experimenting face significant difficulties in simultaneously considering selecting deployment locations based on their intuition and experience while collecting high-quality data during fast-evolving severe weather events and ensuring the safety of operators and vehicles. A timely understanding of the local environment is essential for real-time deployment; however,

on-site, in-person assessment requires time and a lot of human and material resources. A framework consisting of geospatial and meteorological information, geographic information systems (GIS), and location analysis and modeling was developed for field use to delineate feasible areas, eliminate inferior solutions, and identify suitable locations. The problem has been formulated as a coverage location problem, and both exact and heuristic algorithms have been utilized and compared to solve it. The findings highlight the potential of quantitative decision-making techniques in improving the effectiveness of mobile radar deployment, thereby advancing real-time severe weather sensing.

4 - Strategic Placement of Drone Platforms in Out-of-Hospital-Cardiac-Arrest: A Case Study for Jefferson County, Louisville, Ky

Monica Gentili, University of Louisville, Louisville, KY, United States

Each year, over 350,000 out-of-hospital cardiac arrests (OHCAs) occur in the U.S., with survival rates as low as 8–10% due to delayed ambulance arrivals and limited use of publicly available Automated External Defibrillators (AEDs). Unmanned Aerial Vehicles (UAVs), or drones, present a transformative solution for rapid AED delivery to OHCA scenes, overcoming barriers like AED accessibility and bystander hesitation. Using Louisville's EMS data, this study identifies priority areas for drone delivery and potential launch sites in Jefferson County, Louisville, KY. These inputs inform an optimization model to determine the minimum number of launch platforms needed to cover a specified percentage of intervention calls within a desired service time. We present results of the optimization model under different scenarios, by varying the priority areas, the number and location of the platforms, and the drone flight parameters.

SC27

Building B Level 2 B201

Advances in Distributionally Robust and Stochastic Optimization

Invited Session

Computing Society

Chair: Yiling Zhang, University of Minnesota, Minneapolis, MN, United States

1 - Managing Shared Mobility Systems with Electric Vehicles Under Correlated Demand Uncertainties

Kai Pan, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, Ziliang Jin, Bin Tian, Yun Fong Lim

Using electric vehicles (EVs) with vehicle-to-grid (V2G) technology in a shared mobility system promotes sustainability but limits vehicle accessibility. This highlights the importance of optimizing the initial EV allocation, which should also consider subsequent operational decisions. The problem is further complicated by correlated uncertainties in trip demands across service regions and time periods without perfect knowledge of these correlations. We propose a two-stage distributionally robust optimization (DRO) model considering ambiguously correlated trip-demand uncertainties. In the first stage, an operator decides the initial vehicle allocation. In the second stage, the operator determines various operational decisions to meet demands over a time horizon. The objective is to minimize the expected cost under a worst-case joint distribution within an ambiguity set based on moment information of the correlated uncertainties. We show a monotonic relationship between the optimal objective value and the trip-demand covariance matrix. We further prioritize trip-demand pairs based on their covariance value or shadow price, enabling us to focus on a subset of demand pairs, which is especially appealing given the operator's limited resources. To improve computational efficiency, we propose tight approximations of the DRO model based on principal component analysis and develop a hybrid algorithm using a temporal decomposition technique. Numerical results based on real data confirm our approach's efficiency. It is crucial to properly incorporate correlation information, which can attain up to 4.66% total cost reduction.

2 - Bi-Parameterized Two-Stage Stochastic Min-Max and Min-Min Integer Programs

Manish Bansal, Virginia Tech., Blacksburg, VA, United States

We introduce two-stage stochastic min-max and min-min integer programs with bi-parameterized recourse (BTSPs), where the first-stage decisions affect both the objective function and the feasible region of the second-stage problem. To solve these programs efficiently, we introduce Lagrangian-integrated L-shaped (L^2) methods, which guarantee exact solutions when the first-stage decisions are pure binary. For mixed-binary first-stage programs, we present a regularization-augmented variant of this method.

Our computational results for a stochastic network interdiction problem show that the L^2 method outperforms a benchmark method, solving all instances in 23 seconds on average, while the benchmark method failed to solve any instance within 3600 seconds. The L^2 method also achieves optimal solutions, on average, 18.4 times faster for a stochastic facility location problem. Furthermore, we show that the L^2 method can effectively address distributionally robust optimization problems with decision-dependent ambiguity sets that may be empty for some first-stage decisions, achieving optimal solutions, on average, 5.3 times faster than existing methods.

3 - Robust System Identification: Finite-Sample Guarantees and Connection to Regularization

Grani Adiweni Hanasusanto, University of Illinois Urbana-Champaign, Urbana, IL, United States, Hyuk Park, Yingying Li

We study the problem of learning nonlinear dynamical systems from a single sample trajectory. While the least squares estimate (LSE) is a standard approach, it often yields poor identification accuracy when the sample size is limited or the model is misspecified. To address these challenges, we propose a robust LSE framework that leverages robust optimization techniques. We show that this framework is equivalent to a regularized LSE with general Schatten-norm penalties. For linear systems, we establish non-asymptotic performance guarantees, demonstrating a near optimal error rate of $O(1/\sqrt{T})$ that avoids the curse of dimensionality. Empirical evaluations on real-world system identification and online control tasks highlight the substantial performance gains of our method over existing approaches.

4 - Contextual Distributionally Robust Optimization Under Streaming Data: An Alternating Optimization Method

Guzin Bayraksan, The Ohio State University, Columbus, OH, United States, Haoyuan Xue, Merve Bodur

We consider data-driven decision-making that incorporates a prediction model within the 1-Wasserstein distributionally robust optimization (DRO) given joint observations of uncertain parameters and covariates using regression residuals in a streaming-data setting. In this setting, additional data become available and allow decisions to adapt to the growing knowledge of the underlying uncertainty. The ambiguity set shrinks as more data is observed. We propose an efficient online optimization method for this streaming-data contextual DRO setting, which iteratively alternates between optimizing the decision and determining the worst-case distribution. We analyze the asymptotic convergence properties of this algorithm and establish dynamic regret bounds to certify the performance of online solutions. Through numerical experiments, we validate our theoretical findings and demonstrate that our approach significantly enhances computational efficiency while maintaining high solution quality under streaming data.

5 - A Primal Perspective on Distributionally Robust Two-Stage Problems with Integer Recourse

Yiling Zhang, University of Minnesota, Minneapolis, MN, United States

: In this talk, we introduce and study a two-stage distributionally two-stage linear problem with integer recourse, where the objective coefficients are random. The random parameters follow the worst-case distribution belonging to a second-order conic representable ambiguity set of probability distributions. We show that the worst-case recourse objective, under various risk measures, can be formulated as a conic program from a primal perspective. This method also provides additional information on the probability of attaining an integer recourse solution, extending the concept of persistency studied in Bertsimas et al. (2006). Unlike the marginal moment sets used in Bertsimas et al. (2006), the second-order conic representable ambiguity sets in our method offer greater flexibility by incorporating more distributional information. Furthermore, this method enables column constraint generation methods for solving two-stage problems with integer recourse.

SC28

Building B Level 2 B202

Recent Trends in Algorithms and Probability Theory

Invited Session

Computing Society

Chair: Ilias Zadik, Yale, New Haven, CT, United States

1 - New Local Search Algorithms in Sparse PCA

Ilias Zadik, Yale, New Haven, CT, United States

Local-search methods are widely used in statistics, but their theoretical foundations are far less developed than those of low-degree polynomials or spectral methods. A striking example arises in sparse tensor principal component analysis (PCA), where prior work uncovered a large “local-computational” gap: many local Markov chain methods perform significantly worse than the best polynomial-time algorithms.

We address this by proposing new local-search procedures that provably close this gap across multiple regimes of the model, including—and extending beyond—those where traditional local methods fail. Our framework covers two main families: (1) standard greedy and randomized greedy algorithms applied to the (regularized) posterior, and (2) new random-threshold variants. In the latter, the randomized greedy algorithm accepts a move only if the Hamiltonian increase exceeds a Gaussian random threshold, rather than simply being positive.

2 - Learning with Systematic Bias and Missing Data

Alvertos Kalavasis, YALE UNIVERSITY, New Haven, CT, United States

In many problems in Econometrics and Statistics, we only have access to imperfect data due to systematic bias in the data collection process. In this talk, we will present a general formulation for learning in the presence of such biasing mechanisms. Under this framework, we will discuss learning algorithms for: (i) parameter estimation for mixture models, (ii) inference under self-selection bias and coarsening, and (iii) estimation of treatment effects in causal inference.

3 - The Broken Sample Problem Revisited: Proof of a Conjecture by Bai-Hsing and High-Dimensional Extensions

Simiao Jiao, Duke University, Durham, NC, United States, Yihong Wu, Jiaming Xu

We revisit the classical broken sample problem: Two samples of i.i.d. data points $\mathbf{X} = \{X_1, \dots, X_n\}$ and $\mathbf{Y} = \{Y_1, \dots, Y_m\}$ are observed without correspondence with $m \leq n$. Under the null hypothesis, $\mathbf{X}$ and $\mathbf{Y}$ are independent. Under the alternative hypothesis, $\mathbf{Y}$ is correlated with a random subsample of $\mathbf{X}$, in the sense that $(X_{\pi(i)}, Y_i)$'s are drawn independently from some bivariate distribution $\pi: [m] \rightarrow [n]$. Originally introduced by DeGroot, Feder, and Goel (1971) to model matching records in census data, this problem has recently gained renewed interest due to its applications in data de-anonymization, data integration, and target tracking. Despite extensive research over the past decades, determining the precise detection threshold has remained an open problem even for equal sample sizes ($m = n$). Assuming m and n grow proportionally, we show that the sharp threshold is given by a spectral and an L_2 condition of the likelihood ratio operator, resolving a conjecture of Bai and Hsing (2005) in the positive. These results are extended to high dimensions and settle the sharp detection thresholds for Gaussian and Bernoulli models.

4 - An Improved Upper Bound for the Euclidean Tsp Constant

Charlie Guan, Northwestern University, Evanston, IL, United States

Consider n points generated uniformly at random in the unit square, and let L_n be the length of their optimal traveling salesman tour. Beardwood, Halton, and Hammersley (1959) showed $L_n / \sqrt{n} \rightarrow \beta$ almost surely as $n \rightarrow \infty$ for some constant β . The exact value of β is unknown but estimated to be approximately 0.71 (Applegate, Bixby, Chvátal, Cook 2011). Beardwood, Halton, and Hammersley further showed lower and upper bounds such that $0.625 \leq \beta \leq 0.92117$. Currently, the best known bounds are $0.6277 \leq \beta \leq 0.90304$, due to Gaudio and Jaillet (2019) and Yu and Carlsson (2023). In this talk, we present an improved upper bound of 0.89851. In all existing methods that obtain the upper bound, the central idea is to divide the unit square into bands of height $h / \sqrt{n}$ and to construct a path traversing all vertices within each band. The additional length required to merge these band-paths into a valid tour is $O(\sqrt{n})$, so the sum of the lengths of the band-paths is an upper bound on $\beta \sqrt{n}$. Our approach allows paths to cross bands and takes advantage of pairs of points in adjacent bands which are close to each other. Numerical integration techniques yield an upper bound of 0.89851, which is the first upper bound to eclipse 0.90, to the best of our knowledge.

5 - The Price of Sparsity: Sufficient Conditions for Sparse Recovery Using Sparse Measurement Matrices

Youssef Chaabouni, MIT, Boston, MA, United States

We consider the problem of recovering the support of a sparse signal using noisy projections. While extensive work has been done on the dense measurement matrix setting, the sparse setting remains less explored. In this work, we establish sufficient conditions on the sample size for successful sparse recovery using sparse measurement matrices. Bringing together our result with previously known necessary conditions, we discover that, in the regime where $\frac{ds}{p} \rightarrow +\infty$, sparse recovery in the sparse setting exhibits a phase transition at an information-theoretic threshold of $n_{\text{INF}}^{\text{SP}} = \Theta(\frac{s \log(p/s)}{\log(ds/p)})$, where p denotes the signal dimension, s the number of non-zero components of the signal, and d the expected number of non-zero components per row of measurement.

SC29

Building B Level 2 B203

Energy System Design and Operations Under Uncertainty

Invited Session

ENRE: Energy

Chair: Jordan Kern, North Carolina State University, Durham, NC, United States

Co-Chair: Benjamin Rachunok, North Carolina State University, Raleigh, NC, United States

1 - Electrified Systems with Resource Uncertainty: Future Flexibility and Resilience

Constance Crozier, Georgia Institute of Technology, Atlanta, GA, United States

This talk will explore the ability of electrified systems to adapt to possible longer interruptions in energy supply. We will consider how potentially volatile electricity prices would change the capital investments that industrial consumers make, and the benefits that this may have for supply chain and energy resilience. Historically, industrial consumers have coped with relatively stable energy prices and certain energy security. This has led to very “lean” supply chains with little flexibility. The deployment of extremely low cost renewable energy means that supply chain operators stand to profit if they have the flexibility to benefit from cheap variable energy. These effects will be demonstrated using a case study of electrified freight transport along the East Coast.

2 - Transmission Cost Allocation via Stochastic Capacity Expansion Planning

Elizabeth Glista, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, United States, Jean-Paul Watson

Due to booming demand for electric power and aging electric infrastructure, new investments are needed in many existing generation and transmission systems. However, interconnection delays and costs are a huge barrier to new generation and utility-scale storage coming online. During the interconnection process, developers are notified

of the associated cost to upgrade transmission and may withdraw their projects at any point. In many U.S. markets, these transmission upgrades are assigned based on a "beneficiaries pay" approach that seeks to assign the upgrades to those who most benefit from them. However, the benefit of a transmission upgrade in a meshed network is hard to measure, and this cost allocation scheme can lead to system-level investment distortions. In this talk, we demonstrate how transmission cost allocation can be included in a stochastic capacity expansion planning (CEP) model that co-optimizes investments in new generation, storage, and transmission. Then we establish conditions for transmission cost allocation that prevent system-level distortion in both energy-only and capacity markets. While considering the set of active generation and storage projects in the existing interconnection queue, we examine different transmission cost allocation methods on a synthetic-but-realistic model of the state of California that has around 8,000 buses. Using future weather projections as a stochastic input to the model, we demonstrate how our method can be used to establish a transmission cost allocation scheme that results in optimal system behavior over a long-term, uncertain planning horizon.

3 - Critical Mineral Supply Chains for Energy System Modeling: Optimizing Multiple Players and Objectives

Erick Jones, University of Texas at Arlington, Arlington, TX, United States

Critical Minerals, which include elements like Lithium, Cobalt, and Rare Earth Elements, are essential for modern energy and defense applications. There are a variety of concerns about various nations ability to acquire these minerals. Furthermore, even if a sufficient supply is obtainable, refining, and component production has been concentrated in a small number of nations; creating supply chain and even geopolitical risks. Yet even with these known problems most energy systems models exclude supply chain challenges and assume that the materials to build identified technologies are not only available but available at the predicted price. In order to address this gap, this work investigates how to model critical mineral supply chains which include multiple players balancing multiple objectives in an optimization framework that leverages advances in game theory, decomposition techniques, and MILP algorithms to model these supply chains, investigate how different nations will compete for market share, and evaluate how this all affects prices. Then, this work will discuss how these findings can be integrated into energy system modeling workflows to improve accuracy.

4 - Modelling Extreme Weather Risks to Daily US Natural Gas Markets Under near-Term Growth in Electric Power and Lng Export Demand

Cameron Lisy, North Carolina State University, Raleigh, NC, United States, Jordan Kern

Recent steady growth in the production of U.S. domestic natural gas has kept up with rising demand in the power sector, and prices have remained stable despite rising geopolitical headwinds from conflict and uncertainty in trade relationships. But this balance of fundamentals may shift in the near term as more LNG export terminals have begun operations and hyperscale datacenter growth encourages the building of more gas turbines by utilities to meet their 24/7 demand for electricity. Simultaneously, natural gas infrastructure vulnerability to extreme winter storms, most recently experienced from weakened polar vortex events in 2019, 2021, 2023, 2024 and 2025 have raised concerns about the potential for disruption at critical times of year when customers rely on natural gas deliveries both directly and to electric power generators for space heating. To understand the impacts these events and new market conditions may have on the price and availability of natural gas to end-use consumers in key regions of the country and how they may influence each other, we utilize our model of daily U.S. production and state-to-state pipeline flows. We evaluate several possible combinations of supply and demand scenarios, along with simulated disruptions to supply from extreme weather, and explore what they may mean for the future of the natural gas market in the U.S.

5 - Incorporating Equity and Fairness into a Mathematical Model for Power Restoration After Natural Hazards

Benjamin Rachunok, North Carolina State University, Raleigh, NC, United States, Ignacio Sepulveda

The power grid is a critical infrastructure of a community, yet the increasing frequency and intensity of extreme weather events have led to longer, more frequent power outages, disproportionately affecting low-income and minority groups as well as rural communities. To enhance recovery after such events, various restoration strategies have been proposed, focusing on cost reduction and critical infrastructure prioritization. However, these strategies often overlook people's needs, highlighting the need to incorporate equity/fairness into power restoration planning. This study addresses the power restoration planning problem for transmission networks considering customer vulnerability as captured by the Social Vulnerability Index of the United States. In this problem, we employ a Mixed-Integer Linear Programming model to dispatch a set of homogeneous repair crews to fix damaged components, combining operational and power flow constraints for transmission networks, and vehicle routing constraints, testing different objective perspectives. Exact and heuristic methods are proposed to solve the model depending on its complexity. The model's solutions are analyzed in order to understand, first, the cost of equity/fairness; second, the reason of the disparity; last, what is the best strategy for each instance.

6 - A Risk-Averse Stochastic Optimization Approach for Reliably Operating Energy Storage Technology Supply Chains

William Yang, Lawrence Livermore National Laboratory, Livermore, CA, United States, Mengyao Yuan, Elizabeth Glista, Michael Grappone, Jean-Paul Watson

There are unique challenges to reliably managing supply chains for energy storage or generation technologies, as well as other power grid resources. Many grid technologies require assembly from multiple raw materials sourced from across the world. Furthermore, raw materials are often processed at locations that are different than the final production stage, which adds more complexity. There is a high degree of uncertainty in the price and demand for grid technologies because of the ever-changing landscape of power grids and the energy sector. Moreover, uncertainty from market and geopolitical factors impacts the ability to ship materials from one location to another and is exacerbated by the highly interconnected nature of grid technology supply chains. To address these challenges, we develop a risk-averse stochastic optimization methodology to provide secure operation of the energy storage technology supply chain. Our model considers not only the raw material flow, but also the intermediate processing and the final production stage, with each stage taking place in different locations. We additionally incorporate stochastic components such as uncertainties caused by unplanned demand, pricing, and trade dynamics. Our model incorporates risk by optimizing the Conditional Value at Risk metric, which allows us to produce more reliable results compared to traditional stochastic optimization methods. We present experimental results that demonstrate our methodology's efficiency and its ability to reliably provide cost-effective solutions. A major consequence of our methodology is that it can allow grid planners to securely obtain necessary technologies to expand grid capabilities.

SC30

Building B Level 2 B204

Optimization Under Uncertainty for Power Systems Resilience Integrating Distributed Energy Resources

Invited Session

ENRE: Electricity

Chair: Tanveer Bhuiyan, The University of Texas at San Antonio, San Antonio, TX, 78249, United States

1 - A Stochastic Game-Theoretic Optimization Approach for Managing Coupled Local Electricity and Thermal Markets with Electric Vehicles and Renewable Sources

Dai Nam Nguyen, The University of Texas at San Antonio, San Antonio, TX, United States, Tanveer Bhuiyan, Sayed Hamid Hosseini Dolatabadi, Yang Chen

With the growing need for sustainable, resilient, and environmentally friendly energy ecosystems—alongside the rapid adoption of electric vehicles (EVs), renewable energy sources (RESs), combined heat and power systems (CHPs), and heat pumps (HPs)—local energy markets (LEMs) present a distinct potential to promote clean energy, enhance grid stability, and lower emissions. However, unlocking the full potential of LEMs requires addressing complex challenges related to market dynamics, EVs and RESs integration, and electricity-thermal energy interdependency. Our study presents a stochastic bi-level model for managing LEMs, where the local transaction center (LTC) and other market participants—load aggregator (LA), pure thermal industrial prosumers (PTIP), and their lower-level prosumers—aim to maximize their profits in a competitive game-theoretic structure. To ensure grid stability and profitability, the LTC, as the leader, designs a robust dynamic pricing strategy under uncertainty, while the LA and PTIP, as followers, determine the energy transaction amounts under uncertainties in generation and consumption. We address several novel aspects of LEMs, including the heterogeneous behavior of entities, uncertainties in renewable generation and load profiles, and electricity-thermal energy coupling. We develop an exact approach based on Karush–Kuhn–Tucker (KKT) and a parallel dynamic cutting plane algorithm (PDCP) leveraging piecewise linear approximation and

Sampling strategy to efficiently solve the model. Results show that the PDGP algorithm outperforms the KKT-based approach. Results also show that considering the electricity-thermal interdependency increases LTC's profit to 127% of its original value.

2 - Data-Driven Distributionally Robust Optimization for Power System Dispatch

Xunhang Sun, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Xiaoyu Cao, Bo Zeng, Miaomiao Li, Xiaohong Guan, Tamer Basar

Due to the highly intermittent, fluctuating, and random nature, the increasing scale of renewable energy poses huge challenges to the economic and secure operation of power systems. In this work, we present a two-stage distributionally robust optimization (DRO) approach for power system dispatch under supply and demand uncertainties. Specifically, we focus on a practical situation in which the probability distributions of random factors are not known exactly, whereas a limited set of empirical data is available. In this setting, a continuous Wasserstein ambiguity set is constructed to support data-driven decision-making. By fully leveraging the specific structure of the worst-case expectation problem, a discrete equivalence of it is proposed from the *primal* perspective. Then, we customize and develop a novel and exact decomposition algorithm under the framework of column-and-constraint generation for the DRO dispatch problem that guarantees finite-step convergence and facilitates parallel computing. As a byproduct, we shed light on the interrelationships of the DRO dispatch model with the traditional ones based on stochastic programming (SP) and robust optimization (RO). Numerical studies on a multi-energy microgrid demonstrate that, compared to the SP model, the DRO model reduces the probabilities of electricity and heat load shedding from 8.7% to 0.3% and from 12.1% to 1.2%, respectively, with only a negligible 0.3% increase in total cost. And, the DRO model exhibits much less conservativeness than the RO one. The significant improvement of our solution algorithm is verified by comparing to the popular algorithms in the literature for two-stage DRO problems.

3 - An Improved Branch-and-Price Algorithm for Routing and Energy Scheduling of Electric School Buses for Disaster Restoration

Sayed Hamid Hosseini Dolatabadi, The University of Texas at San Antonio, San Antonio, TX, United States, Yuchon Dong, Tanveer Bhuiyan, Bo Zeng

Natural disasters threaten the resilience of power systems, causing widespread power outages that disrupt critical loads (e.g., hospitals) and endanger public safety. Leveraging government-controlled electric school buses (ESBs) having large battery capacity and deployment readiness offers a promising solution for faster power restoration to critical loads during disasters while traditional maintenance is underway. We propose an efficient mixed-integer linear programming model for dispatching, routing, and scheduling a heterogeneous fleet of ESBs to satisfy the energy demand of critical isolated loads around disasters with the minimum cost, addressing the following practical aspects: combined transportation and energy scheduling of ESBs, multiple back-and-forth trips of ESBs between isolated loads and charging stations, and spatial-wise coupling among multiple ESB routes. We develop efficient exact and customized heuristic branch-and-price (B&P) algorithms to solve the proposed model for large networks. Numerical results based on a realistic case study of Florida hospitals demonstrate that our proposed exact and heuristic B&P algorithms are computationally 193 and 343 times faster, respectively, than Gurobi. Using network sparsity to incorporate the limitation in shelter-ESB type compatibility demonstrates that the total cost increases by 217% as the network becomes fully sparse compared to fully connected. The effective usable capacity metric reflects that the largest-capacity ESBs are up to 10 times more effective than the least-capacity ESBs in critical load restoration. Additionally, results demonstrate that the required fleet size increases by 429% as the weather changes from normal to adverse.

4 - Improving the Resilience of the Power Grid to Extreme Heat: Prediction, Evaluation and Optimization

Erhan Kutanoglu, The University of Texas at Austin, Austin, TX, United States

Power grid's operational performance increasingly depends on weather, which shows short- and long-term variability stressing the grid in multiple ways. Temperature and heat waves play a significant role in electricity supply and demand as well as the grid's overall capacity to handle the resulting power flows. We study how to model the long-term weather patterns including climate-shifted temperatures at high temporal and spatial resolutions, how to evaluate their effects on the operational performance of the grid, and how to optimize a resilience improving investment strategy. We use two synthetic test grids mimicking Texas' power system in two different sizes and generation mix profiles and measure the benefits of several optimized planning strategies, including deployment of energy storage systems, for the resilience of the grid against extreme heat.

SC31

Building B Level 2 B205

Electricity Market Design and Pricing Strategies

Contributed Session

Contributed

Chair: Carlos Olivios, Universidad Católica del Norte, Angamos 0610, Antofagasta, Chile

1 - Data-Driven Electricity Price Calibration Based on Bayesian Inference

Haolin Yang, Rensselaer Polytechnic Institute, Troy, NY, United States, Siby Jose Plathottam, Kristen Schell, Todd Levin, Zhi Zhou

Electricity prices in real-world markets can vary widely; prices in the New York Independent System Operator's (NYISO) market, for example, can vary by two orders of magnitude: from a mean of 30/MWh up to 4000/MWh. However, simulated price data from electricity system production cost models (PCMs) result in much narrower distributions. This research has developed a new framework to better calibrate simulated price data from PCMs to real-world price data. This framework is based on Bayesian inference and utilizes two different types of Bayesian models to tackle two problems: Bayesian Ridge Regression to capture extreme price spikes and two Dual-head Bayesian Neural Networks (DBNN) to model prices within the normal range. The calibration framework is validated on real-world data from two regional wholesale electricity markets (California Independent System Operator (CAISO) and NYISO). It is shown that the calibrated PCM data much more closely follows the real-world data distributions, achieving closer skewness and kurtosis values and achieving overall improvement in similarity by as much as 73.58%.

2 - Joint Optimization of Pricing Strategies and Charging Infrastructure Configuration in Urban Networks

Alireza Rostami, Michigan State University, East Lansing, MI, United States, Ali Zockaie

The growing adoption of electric vehicles (EVs), supported by advancements in battery technology and reduced operational costs, introduces new planning challenges for cities and utility providers. While EVs are well-suited for daily urban trips, unmanaged charging demand can stress the power grid, especially during peak hours when users typically prefer to charge. These peaks can increase the gap between average and maximum grid load, requiring costly upgrades in generation, and distribution systems. Moreover, since charging infrastructure is designed to minimize queuing and ensure acceptable service, high peak demand can drive up investment costs. One promising solution is implementing demand management strategies, particularly through dynamic electricity pricing. By shifting charging demand to off-peak hours and less congested locations, pricing mechanisms can improve grid efficiency and reduce infrastructure costs. However, poorly designed pricing schemes can lead to increased detour times, higher user costs, or even lower charger utilization. Most prior work optimizes pricing assuming a fixed charging network, overlooking how pricing impacts charging infrastructure needs. In reality, pricing and infrastructure decisions are tightly linked and should be optimized together. This study develops a large-scale framework that jointly determines station locations, charger counts, and spatiotemporal pricing. It integrates discrete choice models to capture users' responsiveness to pricing, detour time, and SOC, and leverages the POLARIS activity-based model to simulate realistic travel behavior and charging needs. Applied to the Chicago regional network, the framework aims to reduce investment costs, improve user experience, and guide policy design for smarter EV infrastructure planning.

3 - "Credit Scores" for Modern Electric Grid Resources

Trishant Roy, Arizona State University, Tempe, AZ, United States, Kory Hedman

The electric grid is undergoing a fundamental transition from conventional thermal resources to emerging technologies like renewable energy, electric storage, and demand-side management. This brings operational challenges in managing the intermittency and uncertainty characterized by the newer resources. Motivated by standardized financial credit scores, this study aims to propose an analogous scoring mechanism for grid resources to quantify the performance risk associated with fulfillment of the resource's market bid/position. Statistical tools like logistic regression and k-nearest neighbors are being leveraged to develop the framework. The proposed scoring mechanism will allow Independent System Operators (ISOs; entities responsible for ensuring reliable and efficient operation of the electric grid) to assess resources with different characteristics in a unified framework. For example: a controllable solar farm with a strong record of fulfilling its obligations might receive

a score of 750 whereas a hydro-based powerplant with a previous history of outages/disconnections might receive a score of 550. The higher scoring resource (i.e., the controllable solar farm in this example) will be cleared in the electricity market. Hence, this framework allows ISOs to better assess the short-term resource mix of the grid and to maintain reliability. The presentation will delve deeper into the scoring mechanism (dataset creation, features utilized, models tested, and challenges faced).

4 - Improved Price-Taker Optimization of Integrated Energy Systems Using Data-Driven Surrogate Models

Xinhe Chen, University of Notre Dame, South Bend, IN, United States, Daniel Laky, Kay Lu, Alexander Dowling

Integrated energy systems (IESs) combine various energy technologies to enhance operational flexibility and efficiency. Optimization plays a key role in their techno-economic evaluation, but existing models face a trade-off between accuracy and computational cost. While simplified models ignore grid dynamics, rigorous production cost models (PCMs) are computationally intensive due to complex grid constraints and unit commitment problems.

Price-taker (PT) models offer a middle ground by treating electricity prices as fixed inputs. However, they fail to account for how IES decisions may influence market prices, leading to inaccuracies. To address this, we propose a novel optimization framework that integrates data-driven surrogate models into the PT formulation, enabling more accurate representation of market participation effects with a lower computational burden than PCMs.

Our method trains surrogate models on PCM simulation data to capture how IES design and operation affect electricity prices. These surrogates are embedded in the PT optimization to enhance market interaction realism without needing full grid models. We apply this approach to a nuclear power plant retrofitted with a proton exchange membrane (PEM) electrolyzer for hydrogen co-production. The electrolyzer enables flexible operation, allowing the plant to respond to market price fluctuations. Results show that the surrogate-assisted PT model captures key market interactions, improving design and bidding strategies.

This work bridges the gap between computational efficiency and modeling accuracy in IES optimization, offering a scalable solution for market-aware system design.

5 - Forward Trading, Demand Volatility, and Allocative Efficiency

Sebastien Mittraille, Toulouse Business School, Toulouse, France, Henry Thille

We analyze quantity competition between n firms with heterogeneous but constant unit costs of production which face an uncertain demand. Firms can sell forward before demand realizes, and then sell or purchase spot once demand is revealed. The possibility to sell forward ahead of risk and to buy or sell spot once demand is known is common in electricity and gas wholesale markets or in commodity markets (e.g. minerals, fossil fuels, or semi-conductors). Marginal cost heterogeneity corresponds to the situation in which firms operate different technologies of production (e.g. coal, gas, or nuclear electricity generation, mineral resource extraction from different deposits, or cost heterogeneous semi-conductor foundries). Demand volatility can result from the injection of renewable electricity (solar or wind) on the grid whose production is weather dependent, or from geopolitical risks. We show that demand uncertainty partially reverses the welfare gains associated with forward sales in a deterministic setting. Inefficient firms purchase spot to meet their forward commitment in low demand states, and use the forward market to ensure to sell at the expected spot price. Efficient firms sell less on the spot market in low demand states, and reduce forward trading to not overcommit their aggregate sales. Demand volatility reduces the incentives of efficient firms to sell forward, which allows the inefficient firms to increase their forward market share to the detriment of efficient firms. The equilibrium in pure strategies exists in which the number of operative firms vary with demand volatility (holding expected demand fixed). Concentration index decrease.

6 - Mechanism Design in Sector Coupling Communities Using a Game Theory Approach

Yuin Moon, POSTECH, Pohang, Korea, Republic of, Seokwoo Kim, Dong Gu Choi

This study proposes a fair price calculation and settlement mechanism based on the two-step Shapley value for sector coupling energy communities where diverse energy flows such as distributed energy, hydrogen, and heat are linked. In the first step, a coalition is formed through cooperation between stakeholders who have distributed energy resources such as solar and wind power and facilities such as energy storage devices, electrolyzers, hydrogen storage devices, and heat pumps. The value function is calculated based on the energy system model of the formed coalition. In the second step, the Shapley value is calculated based on the derived coalition and value function to perform profit distribution among the coalitions. To evaluate the stability of the proposed settlement mechanism, we examine whether the resulting allocation resides within the core, thereby ensuring that no subset of participants has an incentive to deviate from the community. If the allocation lies outside the core, a subsidy allocation mechanism is proposed to support continued community participation and enhance social welfare from the repeated game perspective.

7 - Strategic Bidding in Electricity Markets Using Metaheuristics

Carlos Olivios, Universidad Católica del Norte, Antofagasta, Chile, Pamela Bustamante

In modern electricity markets, generation companies compete by submitting bids that specify how much energy they are willing to produce and at what price. This process, known as bidding, plays a key role in how electricity is scheduled and priced. Moreover, this decision is complex because companies do not know what their competitors will do in advance. To address this problem, we propose a mathematical model that supports energy producers in creating better bidding strategies. Our model uses ideas from game theory to represent how companies compete in the market. To solve this model, we apply a metaheuristic algorithm. This allows us to find good solutions faster than traditional methods, especially when the problem becomes large or includes uncertainty. Preliminary results of the algorithm tested using synthetic data will be presented.

SC32

Building B Level 2 B206

Coupling of Supply Chains and Energy Systems with Variable Renewable Energy

Invited Session

ENRE: Environment and Sustainability

Chair: Constance Crozier, Georgia Tech ISyE, Atlanta, GA, United States

Co-Chair: Elnaz Kabir, Texas A&M University, College Station, TX, United States

1 - A Data-Intensive Framework for Understanding Residential Energy Needs

Lauren Janicke, Carnegie Mellon University, Pittsburgh, PA, United States, Destenie Nock

Following the recent cuts to the U.S. Department of Health and Human Services and its Low Income Home Energy Assistance Program (LIHEAP), understanding the state of residential energy vulnerability in the U.S. has become even more critical. Past metrics like energy burden and the energy equity gap are valuable; however, they ignore housing characteristics and offer no baseline for residential energy expectations. Past research has focused on developing multi-tier frameworks to measure energy access levels for low- and middle-income countries. We seek to estimate thresholds for the expected energy required to maintain various levels of thermal comfort and safety assuming accessible grid connectivity. Thus, we establish a multi-tier framework focused on understanding energy vulnerability and affordability by using housing characteristics and smart meter data. First, we compare various methods for parameterizing residential energy use from a practical and statistical perspective. Then, we will develop an energy comfort framework using household survey and smart meter data from Arizona's Salt River Project and explore the sensitivity of the framework to

housing characteristics using state-level data from the Residential Energy Consumption Survey. Initial work finds that using a lagged dependent variable in the regression improves model fit the most. This work aims to provide utilities and governments a tool to easily analyze energy vulnerability in their region by incorporating metrics that account for multiple influencing factors.

2 - Toward Net-Zero: a Case Study on Bioenergy System Optimization

Joon Moon, Princeton University, Princeton, NJ, United States, Hongxi Luo, Eric Larson

Decarbonization pathways require energy system models that capture spatial, technological, and temporal complexity, especially for sectors with significant land-use implications like bioenergy. This study introduces a high-resolution supply chain optimization framework designed to identify cost-effective strategies for bioenergy production, conversion, and distribution. The model employs mixed-integer linear programming-based framework to determine optimal infrastructure placement while accounting for economic, environmental, and land-use constraints at a fine geographic scale. It operates in a sequential, myopic fashion, aligning with national-level system optimization by translating aggregated results into localized, actionable strategies. Detailed techno-economic and resource data were compiled across thousands of subnational regions and tested in a key biofuel-producing area under four forward-looking scenarios, including variations in demand growth, electrification trajectories, and environmental protection levels. Findings highlight how infrastructure coordination can support bioenergy scale-up while managing trade-offs in cost, land use, and emissions. This framework offers both a methodological advancement for spatial downscaling in integrated assessment models and a decision-support tool for regional energy planning.

3 - Net-Zero Emissions Heat

Jacqueline Dowling, Stanford University, Stanford, CA, United States, Dimitri Saad, Alicia Wongel, Inês Azevedo, Kenneth Caldeira, Adam Brandt, Steven Davis

Heating accounts for about 40% of U.S. and global CO₂ emissions. We estimate the costs to decarbonize specific industrial and building heating demands in the U.S., and how these costs scale with heat-related carbon emissions avoided for different industrial processes and geographic locations. Our results identify and quantify the most and least challenging heating to decarbonize, target energy infrastructure upgrade priorities, and guide clean heat technology investment and associated research. Local carbon abatement cost data that matches local U.S. decision-making will help advance clean heat deployment.

4 - Implementing Energy Transition Speed Limits in Capacity Expansion Planning

Emil Dimanchev, ZERO Lab, Princeton University, Princeton, NJ, United States, Dominic Davis, Aniruddh Mohan, Eric Larson, Chris Greig, Jesse Jenkins

Net-zero emissions goals imply a transformation of the energy system of an unprecedented speed, scale, and complexity. Energy system optimization models are commonly used to inform decision makers on technological pathways for achieving these goals. However, such tools simplify the ways in which the speed of energy transitions is influenced by constraints on technological development and deployment. We construct a set of constraints, which we refer to as speed limits, that have a meaningful impact on the speed of technology deployment, including: capital discipline processes, supply chain inertia, and endogenous technological learning. We demonstrate the implementation of these speed limits in a state-of-the-art, scalable, open-source energy system optimization model. Our analysis then explores their impact on the U.S. power sector transition and what they imply for the relative merits of different technological pathways. Overall, this work sheds light on avenues for improving the representation of transition speed in energy system optimization modeling.

SC33

Building B Level 2 B207

Data-driven Decision-making in Healthcare

Invited Session

Health Applications Society

Chair: Vishal Ahuja, Southern Methodist University, Dallas, TX, United States

Co-Chair: Yingchao Lan, University of Nebraska-Lincoln, Lincoln, NE, United States

1 - The Ripple Effects of Hospital Closures on Nursing Home Demand, Staffing, and Care Quality

Jong Myeong Lim, Miami Herbert Business School, Coral Gables, FL, United States, Xinyu Liang

Hospital closures have far-reaching consequences for healthcare delivery, yet their impact on nursing homes—a critical component of post-acute and long-term care—remains understudied. This research examines how hospital closures affect nursing home operations, with a focus on the dynamics of demand inflows, care delivery capacity, and quality of care.

2 - Prescriptive Service Rate Model for Healthcare

Sokol Tushe, University of South Florida, Tampa, FL, United States, Yixuan Wang, Donald Lee

Service process plays a central role in operations, but the literature on high resolution models that can be used prescriptively in simulations is relatively sparse. We address this gap by developing a data-driven service rate model that matches observed properties of service times across a range of settings, including healthcare.

3 - Optimizing Lift for a Targeted Set

Shannon Harris, Virginia Commonwealth University, Richmond, VA, United States, Jerrold May, Johnson Moore, Luis Vargas

This paper introduces novel methodologies for binary classification that directly optimize lift at a specified target level. Lift is widely used in practice due to its intuitive interpretation and alignment with many real-world decision-making settings. We propose two linear programming models: one formulated as a mixed-integer program and the other as a bilevel optimization problem. Both are designed to maximize lift at a predefined target level, with the bilevel model additionally generating a ranking of the observations. We apply these models to a no-show prediction problem in outpatient behavioral health clinics, where only a limited number of patients can be contacted with reminders. Compared to logistic regression, our models identify different target groups, thus demonstrating that explicitly optimizing for lift under resource constraints can lead to more effective interventions. These results underscore the value of tailoring classification models to operational realities.

4 - Using Analytics to Build a Reliable Surgery Schedule

Vishal Ahuja, Southern Methodist University, Dallas, TX, United States

Efficient surgery scheduling in operating rooms (OR) is critical for optimizing hospital operations and patient outcomes. The average cost of an unused OR minute is estimated to be \$46. However, scheduling decisions are complex, influenced by factors such as surgery type, provider and room availability, and team familiarity. We develop a machine learning-based predictive model that improves the estimation of surgery durations while also characterizing their uncertainty. Our model leverages historical surgical data and incorporates important features relevant to the surgery. Beyond prediction, we integrate these uncertainty-aware estimates into a robust optimization framework for surgery case scheduling. We demonstrate the effectiveness of our method through a case study using data from a large pediatric hospital. Our work serves as a template for hospitals seeking to enhance operational planning under uncertainty.

SC34

Building B Level 2 B208

Advances in Deterministic Global Optimization

Invited Session

OPT: Global Optimization

Chair: Tristan Gally, FICO, Xpress Optimization, Birmingham, United Kingdom

Co-Chair: Rohit Kannan, Virginia Tech, Blacksburg, VA, United States

1 - A Simple Heuristic to Learn Primal Solutions for AC Optimal Transmission Switching Using Historical Data

Berkay Becu, Georgia Institute of Technology, Atlanta, GA, United States, Santanu Dey

Power systems problems involve solving a series of similar optimization problems that change slightly over time due to fluctuating demand and cost within short time frames. In a transmission network, the AC Optimal Power Flow (AC-OPF) problem determines generator outputs and bus voltages to minimize costs, meet demand, and maintain AC-feasible power flow under physical limitations. The AC Optimal Transmission Switching (AC-OTS) problem extends OPF by reconfiguring network topology through line switching, yielding further cost reduction. Non-convex AC power equations and binary variables for line switching makes the problem non-convex mixed-integer that presents significant computational challenges. This work introduces a heuristic framework that leverages historical data to efficiently generate high-quality primal solutions to the AC-OTS problem. To find primal solutions, we use a sensitivity-based ranking heuristic based on dual values from convex, second-order cone and quadratic, relaxations of AC-OPF as well as the DC-approximation of the problem. These heuristics use the relaxations to acquire possible switching decisions and test them by AC-OPF model which guarantees AC-feasible primal solutions to AC-OTS problem. Since real-world instances have the demand and cost fluctuations in smaller magnitudes, we propose a k-nearest neighbor (kNN) learning approach to identify historically similar cases and reuse their best topologies, found by the heuristics, for new instances. Computational experiments demonstrate that kNN yields near-optimal solutions significantly faster than the expert (heuristics). On average, the expert method took 247.35 seconds with 2.03% cost improvement while kNN algorithm took 10.4 seconds with 1.87% cost improvement.

2 - Learning to Accelerate the Global Optimization of Qcqp: Strong Partitioning and An End-to-End Graph ML-Based Approach

Rohit Kannan, Virginia Tech, Blacksburg, VA, United States, Erin George, Harsha Nagarajan, Deepjyoti Deka

We learn optimal instance-specific heuristics for the global minimization of nonconvex quadratically-constrained quadratic programs (QCQPs). Specifically, we consider partitioning-based convex mixed-integer programming relaxations for nonconvex QCQPs and propose “strong partitioning” to optimally partition variable domains without sacrificing global optimality. Given that solving the strong partitioning problem exactly can be highly challenging, we design a local optimization method using the generalized gradients of the value function in the inner minimization problem. However, solving the strong partitioning problem to local optimality can still be computationally expensive. To address this, we propose a simple AdaBoost regression-based machine learning approximation for homogeneous families of QCQPs. We conduct a detailed computational study on randomly generated QCQP families, including instances of the pooling problem, using the open-source global solver Alpine. Numerical experiments demonstrate that our AdaBoost regression-based approximation of strong partitioning reduces Alpine’s solution time by a factor of 2 to 4.5 on average, with maximum reduction factors ranging from 10 to 200 across different QCQP families. Additionally, we design a novel end-to-end loss function to train a neural network-based policy that directly maps QCQP instance features to corresponding partitioning points, eliminating the need for an expert strong partitioning policy. We also demonstrate how to apply the graph scattering transform to train a single neural network model that generalizes across QCQP families of various sizes. Our results show that this graph scattering-based end-to-end neural network learns a partitioning policy that significantly outperforms the AdaBoost-based approach and generalizes well to larger QCQPs beyond those encountered in training.

3 - Extreme Strong Branching for Qcqp

Dahye Han, Georgia Institute of Technology, Mableton, GA, United States, Santanu Dey, Yang Wang

Since its introduction in 1995, the strong branching technique has proven to be an effective variable selection method in the branch-and-bound algorithm for solving MILPs. While its extension to MINLP is conceptually straightforward, its applicability remains limited. In MILPs, branching creates two child nodes by fixing the branching variable to either 0 or 1. However, branching on a continuous variable requires additional decision on the branching point. Strong branching on a continuous variable typically includes first defining a branching point, and then solving $2n$ relaxations where n is the number of candidate variables to evaluate the improvement by branching on the specific branching point selected a priori.

We propose an “extreme” strong branching, which evaluates multiple branching points for each variable and selects the branching variable and the branching point simultaneously based on the objective value improvement. Despite computational cost, there are two main benefits of this approach: (i) Bound tightening can be done during the branching process. It is similar in the spirit to the FBBT or OBBT implemented in modern global solvers, one key difference is that the bound reduction is a byproduct of the branching process rather than a separate step. (ii) Since the extreme strong branching rule considers improvement in the objective function value, it resembles closely and takes advantage of the strong branching rule that has shown to be successful in the MILP. We show the effectiveness of extreme strong branching on bipartite bilinear problem instances in the structural engineering field.

4 - Recent Advances in the Gurobi Global MINLP Solver

Rodrigo Fuentes, Gurobi Optimization, Plymouth, MN, United States

MILP solver codes have been around for more than 50 years. Their spatial branching counterparts that solve MINLPs to optimality are in an earlier stage of development, making bigger performance improvements easier to accomplish. This talk will discuss performance improvements to the Gurobi MINLP solver as well as support for additional nonconvex nonlinear functions such as the hyperbolic tangent function and the signed power function.

5 - What's New for the Global Solver in FICO Xpress

Tristan Gally, FICO, Birmingham, United Kingdom

This talk discusses the latest updates and improvements to the global optimization capability within FICO Xpress Solver, which allows to solve general mixed-integer nonlinear problems to global optimality. We will discuss the internal workings of the solver, its features, and recent performance improvements, as well as sharing insights about the types of problems that can be solved with global optimizers nowadays.

SC35

Building B Level 2 B209

New Innovations in Public and Societal OR

Invited Session

Public Sector OR

Chair: Weixiao Huang, Worcester Polytechnic Institute, Worcester, MA, 01609, United States

Co-Chair: Carlos Morel Figueroa, Rensselaer Polytechnic Institute, 231 Congress St, Troy, NY, 12180, United States

1 - Optimizing Community Resilience Investments: a Framework for Fair Building Retrofit Planning Under Multi-Hazard Risk

Himadri Sen Gupta, University of Oklahoma, Norman, OK, United States, Andres Gonzalez, Dylan Sanderson

Traditional hazard mitigation planning often struggles to integrate fairness considerations, overlooking disproportionate impacts on vulnerable populations and hindering equitable community resilience outcomes. This research introduces an optimization framework designed to guide fairness-centered planning for community resilience investments through building-level retrofitting. The framework is formulated as a mixed-integer program employing piecewise-linear approximations to operationalize distributive fairness. Specifically, it uses the decomposable Theil index to minimize disparities in mitigation resource allocation across key demographic dimensions (such as housing tenure, race, and income). Simultaneously, the model optimizes for minimized expected economic losses, total building repair times, and population displacement stemming from multi-hazard impacts (earthquake and tsunami). Utilizing hazard impact data derived from the IN-CORE modeling environment, we study the trade-offs between the different objectives for the coastal community of Seaside, Oregon, under 500-year and 1000-year hazard event scenarios and across multiple budget levels

(\$40M, \$80M, \$120M). The results quantitatively demonstrate significant, non-linear trade-offs between minimizing allocation unfairness and achieving efficiency goals, highlighting that substantial budget increases markedly improve the potential for simultaneous gains in fairness and efficiency, but fundamental conflicts often remain, particularly the high marginal cost in efficiency required to achieve the absolute lowest levels of unfairness. This decision-support tool provides planners with a transparent, data-driven method to explicitly evaluate the implications of different mitigation strategies, visualize complex fairness-efficiency trade-offs, and make more informed, evidence-based decisions that enhance community resilience while actively addressing potential disparate impacts.

2 - Optimizing School Bus Pickups with Third-Party Coordination and Delay Prediction

Neha ., Tepper School of Business, Pittsburgh, PA, United States, Peter Zhang, Holly Wiberg

Parents and policymakers throughout the United States report frequent school bus delays and no-shows. Data from New York City (NYC) shows that traffic is responsible for 63% of these delays, followed by aging buses, weather, and driver shortages. In response, we study the school bus routing problem in the presence of traffic and weather-related delays. Specifically, we investigate the benefits of integrating third-party hired drivers alongside regular drivers to mitigate the impact of these delays on the school bus network. We use machine learning models to predict delays by combining historical school bus delay data with traffic and weather information. These delay predictions inform an optimization model, which allocates third-party drivers—at an increased cost—to support delayed routes in the school bus network. This data-driven approach enables schools to minimize disruptions by strategically deploying third-party drivers, ensuring reliable bus operations during delays.

Our findings show that employing third-party services is effective when policymakers prioritize total expected student travel time. However, these benefits are sensitive to the relative weight of student experience and operational costs. The proposed approach is validated through numerical experiments using NYC delay data, highlighting its potential as a tactical tool to manage school bus operations in potential delays.

3 - Impact of Service Operations on Senior-Volunteer Satisfaction in a Non-Profit Organization

Shikha Safaya, Scheller College of Business, Georgia Tech, Atlanta, GA, United States, Basak Kalkanci, Ravi Subramanian

Nonprofit organizations (NPOs) play an essential role in enhancing societal welfare by addressing gaps left by insufficient governmental resources or policies. Predominantly relying on volunteers to deliver critical services such as food and shelter, NPOs significantly contribute to socioeconomic well-being. However, effectively managing volunteers is complex, as NPOs frequently face challenges like volunteer retention without financial incentives, limited scheduling control, and inconsistent participation. We partner with a nonprofit providing meals and personal interactions through volunteers to elderly individuals experiencing limited mobility and food insecurity. Our research investigates how different elements of service design contribute to sustained volunteer engagement and service quality. Additionally, we empirically identify operational characteristics that drive stakeholder satisfaction, including beneficiaries (seniors), volunteers, and donors.

A key finding of our study is that incorporating the service preferences of volunteers and beneficiaries significantly reduces gaps in volunteer engagement and social isolation among seniors. Furthermore, we explore the tradeoff between incorporating these preferences during volunteer allocation to beneficiary demand and managing the mismatch between volunteer supply and beneficiary demand. Our analysis provides prescriptive insights for NPOs on effectively balancing cost considerations with service quality and stakeholder welfare. Specifically, we demonstrate that the volunteers' utility gap between preferred and non-preferred assignments significantly influences their participation decisions. Thus, initiatives such as targeted volunteer training and recognition programs can effectively bridge this gap, enhancing overall volunteer satisfaction and retention.

4 - Cohort Formation for Resource Sharing in the Nonprofit Sector

Weixiao Huang, Worcester Polytechnic Institute, Worcester, MA, United States, Faidra Monachou, Andrew Trapp

Nonprofit organizations (NPOs) increasingly face challenges in accessing physical assets, as well as skills and talent critical to their missions. While collaboration among locally connected, mutually known NPOs offers a partial remedy to resource scarcity, forming effective and sustainable resource-sharing cohorts beyond existing neighborhood bonds, especially with attention to customization, has yet to be effectively addressed. This study proposes a data-informed cohort formation framework to facilitate efficient resource sharing in the nonprofit sector. Our approach integrates empirical experiments and preference elicitation to understand NPO partnership inclinations. These insights then inform the development of partitioning models that generate optimized, preference-aligned cohorts. We evaluate the framework's effectiveness and efficiency through real-world deployments, measuring both the quantity and quality of resource exchanges over time. Our work contributes to the advancement of innovative operations research for social good and offers actionable pipelines to enhance collective impact within the nonprofit sector.

5 - Optimizing a New Metric for Partisan Fairness of Political Districting Plans

Brendan Ruskey, Lehigh University Department of Industrial Engineering, Bethlehem, PA, United States

In this presentation, we describe a new metric for evaluating the partisan fairness of a political districting plan, and explain how this metric can be optimized to generate legal, unbiased districting plans. The new metric tracks how closely the two-party vote-seat curve of a districting plan adheres to an "ideal" vote-seat curve, with more likely shifts in voter partisanship being weighed more heavily. Plans that perform well under our metric satisfy an intuitive definition of partisan fairness. In addition, our metric has clear and notable advantages over existing partisan fairness metrics. We explain how the proposed metric can be used in a mixed-integer program to optimize the partisan fairness of a districting plan, and how a given plan can be heuristically improved to improve its partisan fairness. We present experimental results suggesting that our optimization models can be used to generate compact, fair districting plans that satisfy limits on county splits and Voting Rights Act requirements.

SC36

Building B Level 2 B210

First-order Methods for Function-constrained Optimization and VI Problems

Invited Session

OPT: Nonlinear Optimization

Chair: Digvijay Boob, Southern Methodist University, Dallas, TX, United States

1 - Uniformly Optimal and Parameter-Free First-Order Methods for Convex and Function-Constrained

Zhenwei Lin, Purdue University, West Lafayette, IN, United States

This paper presents new first-order methods for achieving optimal oracle complexities in convex optimization with convex function constraints. Oracle complexities are measured by the number of function and gradient evaluations. To achieve this, we develop first-order methods that can utilize computational oracles for solving diagonal quadratic programs in subproblems. For problems where the optimal value f^* is known, such as those in overparameterized models and feasibility problems, we propose an accelerated first-order method that incorporates a modified Polyak step size and Nesterov's momentum. Notably, our method does not require knowledge of smoothness levels, H^* continuity parameter of the gradient, or additional line search, yet achieves the optimal oracle complexity bound of $\mathcal{O}(\frac{1}{\epsilon^2(1+3\rho)})$ under H^* smoothness conditions. When f^* is unknown, we reformulate the problem as finding the root of the optimal value function and develop inexact fixed-point iteration and secant method to compute f^* . These root-finding subproblems are solved inexactly using first-order methods to a specified relative accuracy. We employ the accelerated prox-level (APL) method, which is proven to be uniformly optimal for convex optimization with simple constraints. Our analysis demonstrates that APL-based root-finding also achieves the optimal oracle complexity of $\mathcal{O}(\frac{1}{\epsilon^2(1+3\rho)})$ for convex function-constrained optimization, without requiring knowledge of any problem-specific structures. Through experiments on various tasks, we demonstrate the advantages of our methods over existing approaches in function-constrained optimization.

2 - Regularized Operator Extrapolation Method for Stochastic Bilevel Variational Inequality Problems

Mohammad Khalafi, Southern Methodist University, DALLAS, TX, United States, Digvijay Boob

The bilevel variational inequality (BVI) problem is a general model that captures various optimization problems, including VI-constrained optimization and equilibrium problems with equilibrium constraints (EPECs).

This paper introduces a first-order method for smooth or nonsmooth BVI with stochastic monotone operators at inner and outer levels. Our novel method, called Regularized Operator Extrapolation (ROPE), is a single-loop algorithm that combines Tikhonov's regularization with operator extrapolation. This method needs only one operator evaluation for each operator per iteration and tracks one sequence of iterates. We show that ROPE gives $\mathcal{O}(\epsilon^{\frac{1}{4}})$ complexity in nonsmooth stochastic monotone BVI, where ϵ is the error in the inner and outer levels. Using a mini-batching scheme, we improve the outer level complexity to $\mathcal{O}(\epsilon^{\frac{1}{2}})$ while maintaining the $\mathcal{O}(\epsilon^{\frac{1}{4}})$ complexity in the inner level when the inner level is smooth and stochastic. Moreover, if the inner level is smooth and deterministic, we show complexity of $\mathcal{O}(\epsilon^{\frac{1}{2}})$. Finally, in case the outer level is strongly monotone, we improve to $\mathcal{O}(\epsilon^{\frac{1}{4}})$ for general BVI and $\mathcal{O}(\epsilon^{\frac{1}{2}})$ when the inner level is smooth and deterministic. To our knowledge, this is the first work that investigates nonsmooth stochastic BVI with the best-known convergence guarantees. We verify our theoretical results with numerical experiments.

3 - Solving Variational Inequality Problem Under Generalized Lipschitz Conditions

Digvijay Boob, Southern Methodist University, Dallas, TX, United States

We present efficient methods for solving monotone variational inequality problems when the operator is not uniformly Lipschitz continuous, and the feasible set may be unbounded. In recent years, optimization problems under generalized Lipschitz conditions have gained traction. However, there is limited literature on generalized conditions for VI problems. We introduce a new type of generalization of Lipschitz conditions for the VI problem that includes convex function-constrained optimization and variational inequality problems, average reward MDPs, and general robust optimization with semi-infinite (or robust) function constraints. In our problem setup, a uniform Lipschitz smoothness constant may not hold, especially since we allow the feasible set to be unbounded. Under the assumption that a VI solution exists and after carefully designing the algorithm, we show that our iterates remain bounded and establish state-of-the-art convergence rate guarantees for various cases, including monotone or strongly monotone operators with possibly generalized nonsmoothness. When the operator is stochastic, we show expected convergence guarantees. We also establish that the iterates are bounded in expectation. Our convergence criterion uses a Nesterov-type sufficiently large radius ball that contains the VI solution. However, unlike Nesterov, we do not need to know this radius to implement our methods for the abovementioned cases.

4 - Variance-Reduced First-Order Method for Convex Function Constrained Optimization with Finite-Sum Objective and Constraints

Yasin Gharehmohammadi, Southern Methodist University, Dallas, TX, United States

This paper investigates the function-constrained convex optimization problem where the objective function $\frac{1}{n} \sum f_i(x)$ has a finite-sum form and function constraints $\frac{1}{n} \sum g_j(x)$ also possibly finite-sum functions. We present a novel Variance-reduced Constraint Extrapolation Method (VR-ConEx), which is motivated by the ConEx method [?] developed for stochastic convex programs. We prove VR-ConEx can obtain reduced gradient complexity utilizing variance reduction compared to the ConEx method. In particular, for the semi-stochastic case, i.e., $n \ll 1$, the VR-ConEx method converges at the rate of $\mathcal{O}(1/K\sqrt{m})$ and $\mathcal{O}(1/K^2)$ for constant and non-constant step size policy, respectively. This gives $\mathcal{O}(m/\sqrt{\epsilon})$ and $\mathcal{O}(m/\sqrt{\epsilon} + 1/\epsilon^{1.5})$ gradient complexity for the constant and non-constant step-size policy, respectively. For the fully-stochastic case, we have proved that the Lagrangian norm is bounded by some constant and then improved the gradient computation. The convergence rate for the constant step size for fully stochastic case is $\mathcal{O}(1/K\sqrt{\max\{m,n\}})$ with $\mathcal{O}(\{m+n\}\epsilon/\max\{n,m\} + \sqrt{\max\{n,m\}}\epsilon)$ gradient computations. For non-constant step size the convergence rate is $\mathcal{O}(1/K^2)$ with $\mathcal{O}(\{m+n\}/\sqrt{\epsilon} + 1/\epsilon^{1.5})$ gradient computations.

5 - Improved Analysis for Large-Scale Difference-of-Convex Algorithms

Anh Duc Nguyen, National University of Singapore, Singapore, Singapore, Alp Yurtsever, Sra Suvrit, Kim-Chuan Toh

We propose and analyze DCA-PAGE, a novel algorithm that integrates the difference-of-convex algorithm (DCA) with the Probabilistic Gradient Estimator (PAGE) to solve structured nonsmooth difference-of-convex programs. In the finite-sum setting, our method achieves a gradient computation complexity of $\mathcal{O}(N + \sqrt{N}/\epsilon)$, surpassing the previous best-known complexity of $\mathcal{O}(N + N^{2/3}/\epsilon)$ for stochastic variance reduced (SVR) DCA methods. Furthermore, DCA-PAGE readily extends to online settings with a similar optimal gradient computation complexity $\mathcal{O}(N + \sqrt{N}/\epsilon)$, a significant advantage over existing SVR DCA approaches that only works for finite sum setting. We further refine our analysis with a gap function, which enables us to obtain comparable convergence guarantees under milder assumptions.

SC37

Building B Level 2 B211

Recent Advances in Nonsmooth and Stochastic Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Ying Cui, University of California, Berkeley, Berkeley, CA, United States

1 - Asymptotic Normality and Optimality of Stochastic Linearized Augmented Lagrangian Method

Jiajin Li, The University of British Columbia, Vancouver, BC, Canada

In their seminal work, Polyak and Juditsky showed that stochastic approximation algorithms for solving smooth equations enjoy a central limit theorem and attain asymptotic-covariance optimality in the local minimax sense of Hájek and Le Cam. More recently, these guarantees have been extended to stochastic nonlinear programming via two well-studied primal methods: (i) stochastic projected gradient descent, which achieves asymptotic optimality but relies on expensive projections and does not identify the active constraint manifold in finite time; and (ii) dual averaging, which identifies the active manifold yet is provably optimal only under affine constraints. To date, no algorithm simultaneously offers projection-free updates, finite-time manifold identification, and asymptotic optimality for general nonlinear constraints. In this paper, we take the first step toward closing this gap by proving that the Stochastic Linearized Augmented Lagrangian Method (SLALM) combines all of these advantages. To that end, we develop two key technical results: (i) Almost-sure convergence of the full primal-dual iterate (ii) Local last-iterate convergence for both primal and dual variables. These results are not only key to establishing the asymptotic normality of SLALM but also provide a proof framework of independent interest for future stochastic primal-dual algorithms.

2 - A Descent-Oriented Subgradient Method for Nonsmooth Optimization

Ying Cui, University of California, Berkeley, Berkeley, CA, United States

Stochastic optimization shares the same fundamental difficulty as nonsmooth optimization, that is, sampled gradients do not vary continuously along the iterates, making the design of convergent descent methods based on merely first-order information particularly challenging. Classical methods in nonsmooth optimization such as bundle techniques and gradient sampling attempt to overcome this by aggregating subgradients from nearby points, but these approaches are often complex and may lack deterministic guarantees. In this talk, I first present a unifying principle behind these methods and introduce a general framework for provably convergent descent algorithms under solely first-order oracle information. Central to this framework is a simple yet powerful technique called subgradient regularization, which generates stable descent directions using only function values and subgradients at a single point, eliminating the need for neighborhood sampling. This framework captures a broad class of nonsmooth marginal functions, including pointwise maxima and minima of smooth functions. Notably, when applied to composite functions, our approach recovers the classical prox-linear method and offers a new dual perspective. I will conclude with numerical experiments showcasing the method's robustness on challenging nonsmooth problems, such as Nesterov's Chebyshev-Rosenbrock function.

3 - Uniformly Optimal Gradient Minimization Using Higher Order Methods

Yao Ji, Georgia Institute of Technology, Atlanta, GA, United States, Guanghui Lan

We propose a novel uniformly optimal framework for minimizing functions that are p -times differentiable with Lipschitz continuous p -th derivatives, targeting approximate solutions with small (sub)gradient norms. The methods achieve the best-known complexity guarantees across convex, uniformly convex, and nonconvex

settings. In the convex case, the number of tensor method iterations required to reach a point with gradient norm at most ϵ is bounded by $\mathcal{O}\left(\frac{L_{p+1} - x^*}{\epsilon^{2/(3p+1)}}\right)$, where L_{p+1} is the p -th order Lipschitz smoothness parameter and x^* is any optimal solution, matching the known lower bound. For uniformly convex functions with parameter σ_{p+1} , the complexity improves to $\mathcal{O}\left(\frac{L_{p+1}}{\sigma_{p+1}^{2/(3p+1)}} \log\left(\frac{f(x_0) - f(x^*)}{\epsilon}\right)\right)$. In the nonconvex case, we introduce a lower curvature property to characterize the degree of nonconvexity and obtain the bound $\mathcal{O}\left(\frac{L_{p+1}^{2/(3p+1)}}{\ell_{p+1}^{1/p - 2/(3p+1)}} \left[f(x_0) - f(x^*)\right] \epsilon^{(p+1)/p}\right)$, where ℓ_{p+1} is the p -th order lower curvature parameter. This improves upon the existing lower bound in terms of dependence on L_{p+1} .

4 - Recent Advances on First-Order Methods for Semidefinite Programming: Rank Identification and Local Linear Convergence

Xin Jiang, Cornell University, Ithaca, NY, United States, Shucheng Kang, Heng Yang

We investigate the performance of the Alternating Direction Method of Multipliers (ADMM) when applied to Semidefinite Programming (SDP). A longstanding belief suggests that ADMM is only capable of solving SDPs to moderate accuracy, primarily due to its sublinear worst-case complexity and empirical observations of slow convergence. We challenge this notion by introducing a new sufficient condition for local linear convergence: as long as the converged primal–dual optimal solutions satisfy strict complementarity, ADMM attains local linear convergence, independent of nondegeneracy conditions. Our proof is based on a direct local linearization of the ADMM operator and a refined error bound for the projection onto the positive semidefinite cone, improving previous bounds and revealing the anisotropic nature of projection residuals. Extensive numerical experiments confirm the significance of our theoretical results, demonstrating that ADMM achieves local linear convergence and computes high-accuracy solutions in a variety of SDP instances, including those where nondegeneracy fails. Furthermore, we identify cases where ADMM struggles, linking these difficulties with near violations of strict complementarity—a phenomenon that parallels recent findings in linear programming.

5 - Analysis of A Class of Heaviside Composite Minimization Problems

Shaoning Han, National University of Singapore, Singapore, Singapore, Ying Cui, Jong-Shi Pang

The minimization of nonclosed functions is a difficult topic that has been minimally studied. Among such functions is a Heaviside composite function that is the composition of a Heaviside function with a possibly nonsmooth multivariate function. Unifying a statistical estimation problem with hierarchical selection of variables and a sample average approximation of composite chance constrained stochastic programs, a Heaviside composite optimization problem is one whose objective and constraints are defined by sums of possibly nonlinear multiples of such composite functions. In this talk, we first present difficulties of tackling such problems using existing optimization techniques. Then we establish stationarity conditions for the problem, and introduce a sufficient condition, termed local convex-like property, under which the proposed stationary point is a local minimizer. Finally, we briefly discuss solution strategies via lifting and reformulation techniques.

SC38

Building B Level 2 B212

Optimization Society Award Session I

Award Session

Optimization

Chair: Jeff Linderoth, University of Wisconsin-Madison, Madison, WI, United States

1 - Khachiyan Prize

Marc Teboulle, School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel

Over four decades, optimization has evolved from the solid ground of convex analysis to the open challenges of nonconvex and data-driven models. This talk retraces a research journey across these transformations: from convex duality and proximal methods to modern nonconvex and adaptive first-order algorithms, highlighting enduring principles, evolving challenges, and the lasting appeal of optimization in theory and applications.

2 - Farkas Prize

Maryam Fazel, University of Washington, Seattle, WA, United States

Over-parameterization (often due to model's scale) has been observed to drastically improve the optimization landscape in learning models, helping standard training algorithms that fail due to bad local minima of the loss function in exact parameterization, remarkably converge to the ground truth under mild assumptions. In this talk, we consider learning Gaussian Mixture Models (GMMs), a fundamental problem in machine learning, for which the Expectation-Maximization (EM) algorithm and its variant gradient-EM are widely used in practice. When the learning model uses the same number of components (m) as the ground-truth model, prior work has shown EM methods succeed in recovering the ground-truth when $m=2$, but fail to do so when $m>2$. We consider the "over-parameterized" case, where the learning model uses $n>m$ components. We show that gradient-EM converges globally: for a well-separated GMM, with only mild over-parameterization $n=O(m \log m)$, randomly initialized gradient-EM converges to the ground truth at a polynomial rate. The analysis relies on novel techniques to characterize the geometric landscape of the likelihood loss, and is the first global convergence result for EM methods beyond the case of $m=2$. Based on joint work with Mo Zhou, Weihang Xu, and Simon Du (Univ. of Washington).

3 - Balas Prize

Minyi Hong, University of Minnesota, Minneapolis, MN, United States

Professor Hong's work has transformed our understanding of distributed, nonconvex, and bilevel optimization, advancing algorithms that resolve longstanding open questions on ADMM (Alternating Direction Method of Multipliers), establish rigorous complexity results for stochastic bilevel problems, and illuminate the structure of minimax formulations. His research has profoundly shaped modern applications in machine learning, signal processing, and AI alignment, while also inspiring new directions across theory, algorithms, and computation. His creativity, depth, and impact embody the spirit of the Balas Prize.

SC39

Building B Level 2 B213

Online/Sequential Allocation for Discrete Optimization Problems

Invited Session

OPT: Integer and Discrete Optimization

Chair: Sebastian Perez Salazar, Rice University, Houston, TX, United States

Co-Chair: Rohan Ghuge, Georgia Institute of Technology, Atlanta, United States

1 - Real-Time Personalization with Simple Transformers

Lin An, Carnegie Mellon University, Pittsburgh, PA, United States, Andrew Li, Vaisnavi Nemala, Gabriel Visotsky

Real-time personalization has advanced significantly in recent years, with platforms utilizing machine learning models to predict user preferences based on rich behavioral data on each individual user. Traditional approaches usually rely on embedding-based machine learning models to capture user preferences, and then reduce the final optimization task to nearest-neighbors, which can be performed extremely fast. However, these models struggle to capture complex user behaviors, which are essential for making accurate recommendations. Transformer-based models, on the other hand, are known for their practical ability to model sequential behaviors, and hence have been intensively used in personalization recently to overcome these limitations. However, optimizing recommendations under transformer-based models is challenging due to their complicated architectures. In this paper, we address this challenge by considering a specific class of transformers, showing its ability to represent complex user preferences, and developing efficient algorithms for real-time personalization. We focus on a particular set of transformers called simple transformers. We show that simple transformers are capable of capturing complex user preferences. We then develop an algorithm that enables fast optimization of recommendation tasks based on simple transformers. Our algorithm achieves near-optimal performance in sub-linear time. Finally, we demonstrate the effectiveness of our approach through an empirical

study on datasets from Spotify and Trivago. Our experiment results show that (1) simple transformers can model/predict user preferences substantially more accurately than non-transformer models and nearly as accurately as more complex transformers, and (2) our algorithm completes simple-transformer-based recommendation tasks quickly and effectively.

2 - Sequential Allocation of Perishable Social Goods: Notification Design Under Heterogeneous Response

Ari Lerner, Yale University, New Haven, CT, United States

Platforms that coordinate the distribution of perishable goods, such as surplus food or medical donations, often notify recipients rather than allocating directly. In these systems, recipients compete to claim resources upon notification, creating a trade-off between efficiency (ensuring goods are claimed quickly) and equity (ensuring fair access across recipients with different response capabilities). This talk examines how notification strategies can be designed to navigate that trade-off. We develop a model in which goods arrive randomly and must be allocated to one of several agencies before perishing. Agencies differ in responsiveness, and the platform selects a notification strategy that determines when each agency is notified. Under natural conditions on the objective, we show that the problem reduces to finding a deterministic notification order. Simulations show that prioritization can improve equity. However, beyond a critical threshold, the optimal strategy shifts to notifying all recipients at once. These findings offer managerial insights for designing notification systems that balance equity and efficiency in allocating perishable resources.

3 - Free-Order Prophet Matching with Patience Constraints: Improved Approximations via Threshold Policies

Calum MacRury, Columbia University, New York, NY, United States

We study a bipartite online stochastic matching problem where the set of feasible edges is known in advance, but the weight of each edge is drawn from a known distribution and revealed only upon querying. The algorithm may adaptively decide which edges to query in an order of its choosing, subject to *patience constraints* that bound the number of queries per vertex. Upon querying an edge, it must immediately and irrevocably decide whether to include it in the matching.

Computing the optimal algorithm is known to be NP-hard, even when one side consists of a single vertex. On the positive side, Pollner, Roghani, Saberi, and Wajc (EC 2022) give a polynomial-time algorithm with an approximation ratio of 0.426 when vertices on one side have unlimited patience, which drops to 0.395 when both sides have bounded patience. Their result is based on rounding a linear program that cannot be rounded within a factor better than 0.544 in the worst case. In this work, we significantly improve these guarantees to $1 - 1/e$ and 0.58, respectively, for the one-sided and two-sided patience settings. A key ingredient in our work is a new configuration LP that allows for more effective rounding.

Our approach is based on designing *threshold policies*, where each queried edge is accepted if its realized weight exceeds a chosen threshold. This framework also extends our results to more general objectives beyond maximizing expected weight, including *revenue* in the sequential pricing problem.

4 - Distributionally Robust Newsvendor on a Metric

Ayoub Foussoul, Columbia University, New York, NY, United States, Vineet Goyal

We consider a fundamental generalization of the classical newsvendor problem where the seller needs to decide on the inventory of a product jointly for multiple locations on a metric as well as a fulfillment policy to satisfy the uncertain demand that arises sequentially over time after the inventory decisions have been made. To address the distributional ambiguity, we consider a distributionally robust setting where the decision-maker only knows the mean and variance of the demand, and the goal is to make inventory and fulfillment decisions to minimize the worst-case expected inventory and fulfillment cost. We design a near-optimal policy for the problem with theoretical guarantees on its performance. Our policy generalizes the classical solution of Scarf (1957), maintaining its simplicity and interpretability: it identifies a hierarchical set of clusters, assigns a "virtual" underage cost to each cluster, then makes sure that each cluster holds at least the inventory suggested by Scarf's solution if the cluster behaved as a single point with "virtual" underage cost. As demand arrives sequentially, our policy fulfills orders from nearby clusters, minimizing fulfillment costs, while balancing inventory consumption across the clusters to avoid depleting any single one. We show that the policy achieves a poly-logarithmic approximation. To the best of our knowledge, this is the first algorithm with provable performance guarantees. Furthermore, our numerical experiments show that the policy performs well in practice.

5 - Parallel Prophets: Dynamic Cut Minimization on Series-Parallel Graphs

William Frendreiss, Georgia Institute of Technology, Atlanta, GA, United States, Alejandro Toriello

We introduce a sequential decision problem on directed graphs. The decision maker adds vertices v in a directed graph and then observes the weight of the arcs out of v from known distributions. The objective is to minimize the expected total weight of the arcs out of the chosen vertices. We explore the connections to the classical prophet inequality setup of Krenkel & Sucheston; in particular, our model can encode an instance of a minimization optimal stopping problem using a directed path. A natural generalization from path graphs is to look at directed series-parallel graphs with source s and sink t . We prove that finding an optimal policy is $\#P$ -hard on this class of graphs. We study a separable lower bound that decomposes the problem into independent, parallel prophet problems, and a heuristic policy that uses this lower bound to decide when to stop. Roughly speaking, the heuristic's approximation guarantee grows with the number of non-sink nodes with in-degree greater than one and is tight even in the case of parallel paths that merge into a single path.

SC40

Building B Level 2 B214

Digital Twin Method and Applications

Invited Session

Service Science

Chair: Guiping Hu, Oklahoma State University, Stillwater, OK, United States

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

1 - A Digital Twin Based on Noisy Data for Productive Maintenance

Samiksha Aryal, Missouri University of Science and Technology, Rolla, MO, United States, Abhijit Gosavi, Susan Murray

A model based on **renewal reward theory** is developed to determine the optimal **preventive maintenance** interval within a **digital twin** (DT) framework. DTs are increasingly being deployed in smart factories to minimize human intervention in equipment monitoring and maintenance.

Unlike traditional maintenance approaches, where a human operator collects data, fits a distribution offline, and then performs optimization to determine the maintenance schedule, the DT automates the entire process. The DT continuously gathers data from the machine to which it is attached, analyzes the data gathered, and automatically triggers maintenance alerts when the machine is close to failure and needs maintenance.

The **time to failure** (TTF) in this work is modeled via an **empirical distribution** here due to its **self-correcting** nature that accommodates adapting to changing data conditions. In ideal conditions (i.e., noise-free data), the empirical TTF results in an increasing failure rate. However, in this study, the data are assumed to be **noisy**, which is typical of real-world conditions. To address this complexity, this research employs a backpropagation **neural network** to fit the TTF distribution. The neural network results in a **smoothed approximation** of the TTF, which is expected to be **more robust** than directly fitting a basic TTF model to noisy data to determine the optimal maintenance interval.

2 - Uncertainty-Aware Digital Twin Manufacturing: a Q-Learning Approach to Robust Production Scheduling

Syed Hasib Akhter Faruqi, Sam Houston State University, Huntsville, TX, United States, Kazi Masum Billah, Rakib Hasan, Hung-da Wan

Machine degradation, quality control, workforce constraints are some of the complicating factors in production scheduling in manufacturing systems. While digital twin technology offers new possibilities for production optimization, handling uncertainty in real-time scheduling decisions still remains a significant challenge. To address this challenge, we develop an uncertainty-aware RL scheduler that quantifies uncertainty by (i) value uncertainty through Bayesian approximation and (ii) data uncertainty through bootstrap sampling. We utilize contextual information from the production system, i.e., defect rate, machine performance, workforce availability, etc. to formulate the reward function. The framework is implemented within a comprehensive digital twin environment simulating a multi-product manufacturing system with machine degradation, quality control processes, and dynamic order management. Our approach extends traditional Q-learning by maintaining probability distributions over Q-values and using uncertainty-informed action selection strategies. The system's performance is evaluated through comparative analysis of four scheduling approaches: standard Q-learning, uncertainty-aware scheduling, Thompson Sampling, and UCB methods. Simulation shows that Thompson Sampling achieves the most significant improvements with 19.5% higher production throughput (2915.40 vs 2440.20 completed items), 99% reduction in demand failures (0.20 vs 17.60 failures), and 99.7% reduction in demand shortfalls compared to standard Q-learning. The uncertainty-aware scheduler shows 4.2% production improvement and 45% reduction in performance variance, while UCB provides 9.4% throughput gains with 37% faster execution time. Parameter sensitivity analysis reveals that Thompson Sampling with skeptical priors yields optimal performance, while UCB demonstrates robustness across hyper-parameter variations. Our findings indicate that explicit uncertainty quantification in Q-learning significantly enhances the robustness of production scheduling decisions.

SC41

Building B Level 3 B301

Data-Driven Decision Making for Healthcare Operations

Invited Session

Health Applications Society

Chair: Tsai-Hsuan (Angel) Chung, The Wharton School, UPenn, Philadelphia, PA, United States

Co-Chair: Hamsa Bastani, Wharton School, Philadelphia, PA, United States

1 - Preference-Aware Models for Workforce Allocation in Healthcare

Katherine Adams, The University of Texas at San Antonio, San Antonio, TX, United States, Arka Roy, Vishal Ahuja, Justin Boutilier, Yonatan Mintz, Sarang Deo

Effective workforce allocation is critical in settings where resources are limited and where making preference-aware decisions aids the retention of patients or employees. This presentation explores two applications of sequential decision-making models tailored to distinct healthcare environments: community-based diabetes care and anesthesiology department operations.

The first project focuses on appointment allocation for a community health worker (CHW) intervention targeting office workers in low- and middle-income countries. Aimed at diabetes prevention and management through lifestyle education, our goal is to maximize enrollment in the care program. Behavioral health interventions often face high attrition rates due to varying preferences in terms of care frequency. To address this issue, we developed a reinforcement learning model that sequentially assigns appointments to be delivered by CHWs by trading off decisions that maximize present enrollment and gather information to improve future enrollment.

The second project focuses on shift allocation in an anesthesiology department. Concurrency protocols have allowed for the expansion of anesthesiologist capacity by allowing a single provider to supervise multiple rooms if each room is staffed with a qualified nurse or resident. This poses an additional burden on the providers, who must oversee multiple rooms simultaneously. Our goal is to enhance provider satisfaction and retention by accommodating individual preferences in shift types, workdays, and team compositions. We developed an optimization model that ensures coverage while promoting team compatibility and provider well-being. Together, these projects demonstrate how sequential decision-making frameworks can be leveraged to improve healthcare outcomes by incorporating individual preferences.

2 - Harmonizing Safety and Speed: a Human-Algorithm Approach to Enhance the FDA's Medical Device Clearance Policy

Mohammad Zhalechian, Kelley School of Business, Indiana University, Bloomington, IN, United States, Soroush Saghaian, Omar Robles

For a medical device to be approved under the FDA's Premarket Notification 510(k) pathway, the manufacturer can submit a claim indicating that it is as safe and effective as another legally marketed device. This regulatory process has been associated with a high recall rate for devices approved under this pathway, raising important concerns over the approach used by the FDA. We develop a combined human-algorithm methodology to assist the FDA in improving the 510(k) medical device clearance process by reducing the risk of potential recalls and the workload of the FDA. We first develop machine learning methods to estimate the risk of recall for 510(k) medical devices based on the information available at the time of submission. We then propose a data-driven clearance policy that recommends acceptance, rejection, or deferral to the FDA's committees for in-depth evaluation. We conduct an empirical study using a unique, large-scale dataset, and our results show a 31.8% improvement in the recall rate and a 40.5% reduction in the FDA's workload. Our analyses also indicate that implementing our policy could result in significant cost savings.

3 - DA BE ACTIVE: Piloting Dynamic Adaptation

John Silberholz, University of Maryland Smith School of Business, College Park, MD, United States, Hamsa Bastani

Dynamic adaptation is a promising strategy for scaling the impact of behavioral interventions by personalizing interventions based on individuals' needs. Key challenges include designing pilots that create sufficient dynamic variation in intervention strategies, identifying promising sources of heterogeneous treatment effects with limited and noisy data, and evaluating the resulting policies online. We report progress on these fronts on a physical activity intervention motivated by the BE ACTIVE randomized controlled trial, which demonstrated significant and durable step count gains in a high-risk patient population when using gamification or providing financial incentives. Our pilot, which we call the Dynamically Adaptive (DA) BE ACTIVE Pilot, allocates all participants to gamification to encourage step count increases. Participants who fail to meet their step count goals in a given week are randomized to one of three interventions in the following week: study staff outreach, financial incentive, or no additional intervention. Allocation probabilities are dynamically adapted based on accruing study data.

4 - To Batch or Not To Batch: The Impact of Batching Behavior in Emergency Department Prescription Orders

Yong Xia, University of California, San Diego, La Jolla, CA, United States, Danqi Luo, RJ Niewoehner

Emergency departments (EDs) are under constant pressure to balance efficiency and quality of care, yet diagnostic test ordering patterns remain a critical and understudied factor in this equation. We know operational characteristics influence test ordering, where ordering tests early in a visit reduces congestion (Batt and Terwiesch 2017), or where the testing process can alter test-ordering likelihood (Jaeger and Tucker 2019). In recent years, the frequency of diagnostic test orders has increased substantially (O'Sullivan et al. 2018), leading to our primary research question: How do physicians structure their test orders over multiple patients, and what implications does this have for efficiency and patient outcomes?

One understudied aspect of test ordering is batching behavior, where physicians see multiple patients in succession before submitting orders collectively. While prior research has examined when and how frequently tests are ordered (e.g., Ergun-Sahin et al. 2022), little attention has been given to the ordering process itself, specifically, whether batching affects the quantity of tests ordered, resource utilization, or patient outcomes. This study addresses this gap by evaluating the impact of batching on test-ordering patterns, ED operational efficiency, and healthcare costs. We analyze detailed patient-order data from two distinct healthcare settings to identify how batching shapes test utilization. Our econometric approach accounts for confounding factors such as physician workload and ED congestion, ensuring a robust analysis of batching behavior. Our findings provide insight into whether batching should be discouraged as a source of unnecessary test overuse or recognized as an efficiency-enhancing practice in emergency care.

SC42

Building B Level 3 B302

Analytics for Healthcare Operations: Methods and Applications

Invited Session

Health Applications Society

Chair: Fernando Acosta-Perez, University of Wisconsin-Madison, Madison, WI, United States

1 - Interconnected Care: System-Wide Intelligence for Optimizing Hospital Flow

Fernando Acosta-Perez, University of Wisconsin-Madison, Madison, WI, United States, Gabriel Zayas-Caban, Justin Boutilier, Sabrina Adelaine, Frank Liao, Brian Patterson

Coordinating emergency department (ED) admissions is a challenging task for hospital systems, requiring decision makers to balance multiple competing objectives such as minimizing patient waiting times, ensuring appropriate level-of-care placements, and avoiding downstream bed blocking. These challenges are amplified in multisite hospital networks, where increased interconnectivity presents new opportunities for coordinated decision making across facilities. This talk explores how health systems can leverage interhospital connectivity to improve patient flow. We first examine the role of real-time discharge predictions to support proactive admission planning. Using historical discharge data from our partner system, we evaluate strategies for forecasting short-term discharges across multiple hospitals and discuss their implications for operational decision making. Although such models can serve to aid decision makers with complicated decisions, it is not clear how can they transform such predictions into actionable insights. Building on this, we propose a markov decision process framework for inpatient admission control that models the complex trade-offs involved in routing patients across multiple hospitals. Given the high-dimensional nature of this problem, we explore approximate solution strategies that draw insights from the fluid control problem. Through simulation experiments informed by real-world hospital data, we evaluate the potential of coordinated policies that allow interhospital transfers compared to policies that restrict admissions to the arrival hospital. Preliminary results suggest that coordinated strategies may lead to reductions in ED boarding and improve overall patient placement.

2 - Ambulance Allocation for Patient-Centered Care

Eric Stratman, Columbia University, New York, NY, United States, Laura Albert, Justin Boutilier

Emergency Medical Services (EMS) in the United States and similar systems typically utilize a single treatment pathway, transporting all patients to emergency departments (EDs), regardless of their actual care needs or preferences. Recent policy reforms have sought to introduce alternative treatment pathways to divert lower acuity patients from the ED, but operationalizing these options has proven difficult. This paper proposes a patient-centered EMS (PC-EMS) ambulance allocation model that supports multiple care pathways by aligning EMS responses with individual patient needs. We develop a two-stage mixed-integer optimization framework that incorporates multiple dispatch and secondary assignment strategies which enable dynamic resource deployment. The model maximizes appropriate ED diversions while maintaining ambulance availability using a queueing-based availability constraint. We leverage national EMS data and machine learning to estimate dispatcher accuracy and diversion potential. Simulations across diverse geographic regions suggest that agencies can achieve up to 80% of possible ED diversions by equipping only 15–25% of their fleet with diversion-capable units. Adaptive dispatch strategies improve diversion rates by 3.4 to 8.6 times compared to conventional single-unit dispatch. These results demonstrate a critical trade-off between equipment investment and operational coordination, allowing agencies to strategically balance fleet upgrades against the complexity of dispatch protocols based on their specific organizational constraints.

3 - Dynamic Interval Prediction of Patient Waiting TIMES for Cancer-Care Facilities

Yashi Huang, University of Rochester, Rochester, NY, United States, Arik Senderovich, Yaron Shaposhnik

We develop a system for a large cancer institute for predicting in real-time patients' waiting times. The system consists of a collection of machine learning and simulation models that are assembled to make predictions (e.g., different models are used in different cases in different stages of the service process). The system is designed to overcome various challenges such as stochasticity (e.g., unexpected cancellations, and random service and arrival times), heterogeneity (e.g., of patients, providers, and workflow), complex service process (e.g., multiple interconnected stages), and limitations of the available data (e.g., noisy data and changes to the schedule that may not be electronically recorded). At the same time, the system is designed to provide the users explanations about why predictions are made and alert them about potentially inaccurate predictions and delays to the schedule. We report on the performance of our models for the various prediction tasks using a real-world data set.

4 - Fairness and Efficiency of Teledermatology Triage Policies

Le Sun, University of Rochester, Rochester, NY, United States, Yaron Shaposhnik, Vera Tilson

We examine a hybrid teledermatology triage system where certain patients access electronic screening (E-screening) while others go through face-to-face (FTF) screening, with both pathways potentially followed by a FTF treatment. This dual-pathway approach, despite offering operational benefits, creates a tradeoff between system efficiency and equitable access. The disparity manifests because E-screening remains inaccessible to a substantial portion of the population due to technological or institutional barriers. Our research examines how different prioritization policies affect system efficiency and fairness.

5 - Constrained Distributional Reinforcement Learning in Healthcare

Zequan Chen, Dartmouth College, Hanover, NH, United States

The medical system often faces understaffing and time pressure, which can cause patients to miss critical treatment windows. To address this, we propose an automated treatment recommendation system. Existing systems typically rely on expected outcomes to evaluate treatments, recommending the option with the highest expected benefit. However, informative decision-making often requires more than just expectation. Motivated by this limitation, we develop a treatment recommendation framework based on distributional reinforcement learning (DRL), which models the full distribution of potential outcomes for each treatment. This enables personalized recommendations that align with the preferences of both doctors and patients. Furthermore, we incorporate constraints into the DRL process—our constrained DRL algorithm can account for constraints in fairness, financial limitations, and treatment risks. Numerical results demonstrate that our method achieves near-optimal treatment outcomes while respecting these practical constraints and supporting personalized, stakeholder-driven decisions.

SC43

Building B Level 3 B303

Optimization and Analytics for Decision-Making in Healthcare

Invited Session

Health Applications Society

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

Co-Chair: Sun Ju Lee, Bucknell University, Atlanta, GA, United States

1 - A Novel Adaptive Testing Scheme for Multi-Disease Testing

Berk Gorgulu, DeGroote School of Business, McMaster University, Hamilton, ON, Canada, Douglas Bish, Ebru Bish, Merve Bodur

Disease testing plays a critical role in disease mitigation by enabling early detection and timely intervention. However, the high cost of testing—especially in settings requiring screening for multiple diseases—limits its widespread implementation. Two cost-saving strategies are commonly employed: *multiplexing*, where a single assay tests for multiple diseases, and *pooling*, where samples from multiple individuals are tested together. While combining these strategies creates reduction in testing costs, the standard Dorfman testing—using the same multiplex assay for both pooled and individual follow-up tests—can be inefficient, particularly when only a subset of diseases triggers retesting.

This work introduces and analyzes a novel strategy: *adaptive testing*, which leverages laboratory automation to customize retesting assays based on preliminary pooled test results. By adding disease-specific reagents only for relevant diseases during retesting, adaptive testing can reduce costs significantly compared to Dorfman pooling or large multiplexing alone. We develop a rigorous analytical framework for multi-disease testing with flexible retesting and derive structural properties of the optimal solutions. Algorithmic methods are proposed to solve the resulting complex combinatorial problem. Through a case study on 18 respiratory diseases using real data, we show that adaptive pooling can reduce testing costs significantly relative to current best practices.

2 - Data-Driven Optimization of Clinical Treatment Guidelines

Sun Ju Lee, Georgia Institute of Technology, Atlanta, GA, United States, Gian Garcia

Evidence-based guidelines play an important role in informing how chronic diseases should be managed, as these recommendations are widely disseminated and widely implemented. However, they are one-size-fits-most, failing to account for patient-to-patient differences. Personalized medicine has shown significant potential to improve health outcomes over clinical practice guidelines. However, the implementation of personalized medicine may be challenging, resulting in unwanted practice variation and suboptimal patient care. To find the right balance between personalized medicine and clinical guidelines, we propose the treatment guideline design problem. We develop a framework to design optimal treatment guidelines for a population of patients which are each modeled according to their own Markov Decision Process parameters. We propose several exact and heuristic methods to solve the problem, including a mixed-integer linear program formulation and a branch-and-bound clustering algorithm. Our results on a hypertension treatment planning case study demonstrate that with only a small number of subgroups, we can develop clinical practice guidelines which perform almost as well as personalized treatment policies.

3 - Inventory Optimization for Pharmaceutical Fulfillment Networks: a Predictive and Prescriptive Analysis at Amazon Pharmacy

Yinsheng Wang, University of Washington, Seattle, WA, United States, Alexandre Alves, Pierre Bhoorasingh, Cristobal Pais, Xin Tang, Shan Liu

Effective inventory management is critical for all pharmacies, but becomes even more challenging for nationwide digital-forward pharmacies, where timely delivery, regulatory compliance, and diverse product requirements pose significant operational challenges. Amazon Pharmacy continues to expand its services, eliminating pharmacy deserts and expanding sub-same day prescription delivery. This study develops a data-driven inventory optimization framework, integrating predictive and prescriptive analytics at a pharmaceutical fulfillment center. Utilizing de-identified operational data, we first applied automatic statistical distribution fitting to characterize demand patterns accurately. Customer abandonment behavior was explicitly modeled to reflect realistic operational dynamics. Next, we implemented and systematically compared several inventory replenishment policies, including (s, S), Newsvendor, On-Demand, and Safety Stock, through extensive simulation experiments. Empirical results reveal that dynamically selecting replenishment policies tailored to distinct demand profiles significantly outperforms static policies, effectively reducing inventory costs and stockouts. Additionally, critical operational constraints such as cold chain storage capacity were incorporated, demonstrating robustness of the approach under realistic operational conditions. Overall, this study provides critical methodological insights and practical contributions to inventory management strategies specifically tailored for Amazon Pharmacy, leading to potential 18.10% savings of average cost compared to current production baseline.

4 - Machine Learning-Based Risk Stratification for Corneal Transplantation Outcomes in North India: a Retrospective Analysis

Qian Deng, University of Iowa, Iowa city, IA, United States, Nick Street, Simran Sarin, Kanwal Singh Matharu, Neha Kapur

Corneal diseases remain a major contributor to blindness in India. In this study, we retrospectively analyzed the clinical records of 911 patients who underwent corneal transplantation at a leading non-profit tertiary referral eye center in North India between October 2021 and February 2024. Among all evaluated clinical features, Therapeutic Penetrating Keratoplasty (TPK) and infection by *Pythium* species emerged as highly significant predictors of graft failure in classification models. An XGBoost-based predictive model achieved robust stratification into high- and low-risk groups, with Kaplan–Meier survival curves demonstrating significant differentiation (log-rank test p-value = 0.0064). Our findings underscore the potential of machine learning methods, particularly XGBoost, in providing valuable decision support for clinical management and identifying critical factors associated with graft failure following corneal transplantation in India.

5 - Inverse Quadratic Optimization with Application to Cancer Radiotherapy

Archis Gbate, University of Minnesota, Minneapolis, MN, United States

In a typical (forward) optimization problem, a decision-maker uses given values of model parameters to compute the values of decision variables. The goal in inverse optimization (IO) is instead to infer parameters that render given values of decision variables optimal. Most papers on IO utilize duality to impute objective function parameters. A corresponding literature for imputing constraint parameters is essentially non-existent, even for linear programs. The difficulty is that these IO problems include nonconvex bilinear constraints and/or objectives. We will discuss models and solution methods designed to tackle this challenge when the forward problems are quadratic. That is, when the objective is quadratic and the constraints are either linear or quadratic. We will illustrate key ideas using the motivating problem of optimizing treatment plans in intensity modulated radiotherapy for cancer.

SC44

Building B Level 3 B304

Innovations in Adaptive Decision Systems for Healthcare

Invited Session

MSOM: Healthcare

Chair: Michael Lingzhi Li, Harvard Business School, Boston, MA, 02163, United States

1 - Regulating Adaptive Medical Artificial Intelligence: Can Less Oversight Lead to Greater Compliance?

Liang (Leon) Xu, Singapore Management University, Singapore, Singapore, Jiayi Lai, Xin Fang, Tinglong Dai

As of June 2024, the U.S. Food and Drug Administration (FDA) has approved 950 medical artificial intelligence (AI) devices. The current regulatory framework freezes AI algorithms after approval, requiring new submissions for updates to ensure compliance with Good Machine Learning Practices (GMLP). This approach imposes a significant administrative burden, while hindering the ability of AI algorithms to learn from new data. To address these challenges, the FDA has explored a novel pathway known as Predetermined Change Control Plans (PCCP), allowing developers to outline future changes during initial submissions and exempting approved changes from regulatory review. Yet, the impact of this exemption on GMLP compliance remains uncertain. In this paper, we model the strategic interaction between a developer and a regulator in a two-stage game with asymmetric information. The developer may choose to follow or deviate from GMLP in developing and retraining the AI algorithm, whereas the regulator reviews the marketing-clearance application for approval. Our analysis shows that, contrary to intuition, less review can actually lead to greater compliance. This scenario arises, even without considering the administrative burden saved, when (1) auditing capability is moderate and (2) the potential for efficiency improvements through retraining is substantial. Conversely, reclearance is valuable when regulatory review effectively detects noncompliance or when efficacy improvements from retraining are unlikely. We also show adaptive algorithms offer advantages over frozen algorithms in terms of not only improved device efficiency but also greater compliance. Interestingly, these advantages are particularly salient when regulatory oversight has limited ability to detect noncompliance.

2 - Assortment Optimization for Patient-Provider Matching

Naveen Raman, Carnegie Mellon University, Pittsburgh, PA, United States, Holly Wiberg

Rising provider turnover forces healthcare administrators to frequently rematch patients to available providers, which can be cumbersome and labor-intensive. To reduce the burden of rematching, we study algorithms for matching patients and providers through assortment optimization. We develop a patient-provider matching model in which we simultaneously offer each patient a menu of providers, and patients subsequently respond and select providers. By offering assortments upfront, administrators can balance logistical ease and patient autonomy. We study policies for assortment optimization and characterize their performance under different problem settings. We demonstrate that the selection of assortment policy is highly dependent on problem specifics and, in particular, on a patient's willingness to match and the ratio between patients and providers. On real-world data, we show that our best policy can improve match quality by 13% over a greedy solution by tailoring assortment sizes based on patient characteristics. We conclude with recommendations for running a real-world patient-provider matching system inspired by our results.

3 - Optimal Algorithms for Online Scheduling

Anvit Garg, Harvard University, Cambridge, MA, United States, Michael Lingzhi Li

We study an online scheduling problem for a single hospital room, where patients arrive sequentially and request a known length of stay (LOS). Upon each arrival, the decision-maker must accept or reject the patient, earning a LOS-dependent reward only if the patient is admitted. We formulate the problem as a Markov Decision Process (MDP) and analyze the structure of the optimal policy. We show that it admits a compact, transformed-threshold form. Through theoretical analysis and simulation, we explore how this threshold depends on key factors such as the discount rate, arrival rate, and the distribution of LOS. Our results offer insight into interpretable, low-dimensional policies for real-time decision-making in constrained healthcare settings. We also explore extensions to cases where LOS is estimated rather than known.

4 - The Human-AI Dance: Dynamic Adaptive Learning for Optimal AI Guidance

Shirley Huang, Harvard Business School, Boston, MA, United States, Michael Lingzhi Li

As AI systems increasingly assist humans in high-stakes, repeated decisions, it becomes crucial to design optimal collaboration strategies. Most existing approaches assume a static human-AI system, overlooking how users adapt to algorithmic guidance through repeated interactions. We propose a dynamic framework for human-AI collaboration, modeled as a Markov Decision Process (MDP) and selectively provides recommendations to a human whose trust in the AI evolves based on past agreement. Using perturbation analysis, we derive interpretable closed-form approximations that characterize nonlinear tradeoffs between building trust and optimizing short-term outcomes. Our analysis reveals when it is optimal to withhold or even reverse AI recommendations to strategically manage trust and improve long-term performance. Building on these insights, we develop a practical recommendation algorithm that learns the trust threshold online and intervenes only when guidance is expected to improve cumulative outcomes. Simulated and empirical experiments demonstrate that our approach preserves trust and achieves superior long-term rewards compared to always-advise and myopic baselines. These findings offer actionable insights for designing trust-aware AI decision support systems across healthcare, operations, and digital platforms.

5 - Structural Estimation of Treatment Heterogeneity in Sequential Decision-Making Tasks

Ziyuan Zhang, University of Toronto, Toronto, ON, Canada, Ningyuan Chen, Rodrigo Mansur, Vahid Sarhangian

Multi-armed bandit (MAB) tasks are widely used in psychiatry to assess human decision-making behavior, particularly how they navigate the trade-off between exploration and exploitation (E&E). In this work, we propose structural estimation models to estimate the heterogeneous treatment effect of a clinical intervention (e.g., a pharmacological intervention) on the E&E behavior of the participants of a MAB tasks. We consider a randomized experimental design where all participants complete a MAB task twice: once before and once after receiving either an intervention or placebo. To evaluate the treatment effect, we adopt a difference-in-differences (DID) approach that compares pre- and post-treatment periods. We apply the proposed methods to data from a randomized trial involving patients with major depressive disorder (MDD) who completed a four-armed bandit task before and after receiving either a GLP-1R agonist (the active substance in Ozempic) or placebo. The data also includes individual characteristics, such as age, BMI, and depression level, potentially resulting in heterogeneity in treatment responses.

SC45

Building B Level 3 B305

MSOM Student Paper Competition 1

Award Session

Manufacturing and Service Operations Management (MSOM)

Chair: Hummy Song, Wharton, Philadelphia, PA, United States

Co-Chair: Andre Calmon, Scheller College of Business, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Georgina Hall, INSEAD, FONTAINEBLEAU, France

1 - tbd

2 - tbd

3 - tbd

4 - tbs

SC46

Building B Level 3 B306

Risk Management and Strategic Incentives

Invited Session

Decision Analysis Society

Chair: Ruiting Zuo, Financial Technology Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, N/A

1 - Multi-Market Cournot Competition with Investments

Boxuan Li, Columbia University, New York, NY, United States

We develop a two-stage model of multimarket Cournot competition in which firms strategically invest in cost-reducing technologies before competing across multiple interconnected markets. Our model explicitly incorporates endogenous investments into firms' production-cost structures, enabling the analysis of how investment affects firms' competition and welfare in a networked economy. We find that firms' investments are strategic substitutes: the investment of one firm imposes a negative externality on others and reduces their incentive to invest. We show that an inefficiency arises when the adjustment cost is high, pushing large firms to invest more than what is socially optimal. We demonstrate that if adjustment costs are more homogeneous across firms, less central firms benefit from investment which in turn reduces market concentration. Conversely, if cost heterogeneity is higher, more central firms can increase their competitive advantage over smaller firms and increase market concentration. Our analysis suggests that policies offering subsidies to smaller firms to mitigate their higher adjustment costs can effectively prevent excessive market centralization.

2 - A Risk Sensitive Contract-Unified Reinforcement Learning Approach for Option Hedging

Xianhua Peng, Peking University, Shenzhen, China, People's Republic of

We propose a new risk sensitive reinforcement learning approach for the dynamic hedging of options. The approach focuses on the minimization of the tail risk of the final P&L of the seller of an option. Different from most existing reinforcement learning approaches that require a parametric model of the underlying asset, our approach can learn the optimal hedging strategy directly from the historical market data without specifying a parametric model; in addition, the learned optimal hedging strategy is contract-unified, i.e., it applies to different options contracts with different initial underlying prices, strike prices, and maturities. Our approach extends existing reinforcement learning methods by learning the tail risk measures of the final hedging P&L and the optimal hedging strategy at the same time. We carry out comprehensive empirical study to show that, in the out-of-sample tests, the proposed reinforcement learning hedging strategy can obtain statistically significantly lower tail risk and higher mean of the final P&L than delta hedging methods.

3 - Multi-Product Newsvendor with Submodular Demand Restrictions

Linda Punt, Rotterdam School of Management, Rotterdam, Netherlands, Daniel Chen, Gar Goei Loke, Yashar Ghiassi

We study a novel subclass of the multi-product newsvendor model where sales is restricted by a maximal demand that is submodular and random. This setting is prevalent in the inventory and revenue management literature and arises from (i) product indifference, where customers perceive no difference in net utility among a set of alternatives, and (ii) fulfillment deference, where due to indifference, choice of the specific product to fulfill customer demand is deferred to the retailer. We characterize the exact optimal solution, showing that optimal inventory levels should be decided greedily by prioritizing products in decreasing sequence of selling price, while cost only affects the specific inventory levels. Furthermore, these inventory levels are the critical quantiles of some demand distribution, akin to the traditional newsvendor

problem. However, this distribution also captures the demand of other products that the relevant product could fulfill due to product indifference. Our work calls for deeper examination into stochastic optimization over submodular random variables.

4 - Supply Chain Strategy and Risk Hedging

Jingren Zhao, The University of Hong Kong, Hong Kong, Hong Kong, Liao Wang

Push and pull are two common supply chain strategies. In a push system, the retailer (newsvendor) bears all the demand risk, as the supplier "sells to" the retailer. In contrast, a pull system allows the retailer to return unsold units to the supplier, who then assumes the demand risk.

It is well established that, in a risk-neutral setting, pull strategies generally coordinate the supply chain more effectively than push strategies. However, in a risk-averse setting where the parties have differing risk aversion levels, prior literature shows that pull does not necessarily outperform push—this depends on the asymmetry in risk aversion.

In this work, we assume that the party bearing demand risk (the supplier in pull and the retailer in push) can employ risk hedging using financial assets correlated with demand. We then analyze how hedging influences whether push or pull better coordinates the supply chain. Additionally, we examine a scenario where the retailer has a more accurate demand forecast than the supplier and study how risk hedging affects both strategies.

5 - Dynamic Contracting for Patent Screening with AI Techniques

Ruiting Zuo, Financial Technology Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China, People's Republic of, Jussi Keppo, Jiahui Shi

We study incentive contracts within a continuous-time principal-agent framework, where a principal hires an agent to review patent applications. The agent exerts costly efforts in both the quantity and quality of the review process. The principal cannot directly observe the agent's actions but instead monitors a stochastic volume process of examined applications, and a jump process that captures the occurrence of adverse events, such as the erroneous granting of unqualified patents or the rejection of qualified ones. Our model characterizes the structure of the optimal dynamic contract that simultaneously incentivizes the agent to balance examination speed and accuracy under moral hazard. Furthermore, we analyze the interaction between the agent's efforts in quantity and quality across different scenarios and provide a comparative evaluation of human examiners' performance relative to AI-assisted review systems.

SC47

Building B Level 3 B307

Strategic Marketing and Channel Design

Contributed Session

Contributed

Chair: Micky Mukesh Chhabda, University of Connecticut School of Business, 3260 Bloomfield Ln Apt 723, Auburn Hills, MI, 48326, United States

1 - Balancing Control and Cost in Automotive Strategies: a Game-Theoretic Model for Sales and Service Mode Selection

Zhepei Chen, Tianjin University, Tianjin, China, People's Republic of, Zhaofang Mao

With the rise of new energy vehicle manufacturers and increasing competition in the automotive market, the traditional 4S dealership model is undergoing profound transformation. Sales and after-sales services have become two critical sources of profit for car manufacturers, prompting a reevaluation of distribution and service strategies. This study develops a game-theoretical framework to analyze a manufacturer's optimal choice between agency and reselling models for vehicle sales, as well as self-operated versus outsourced models for after-sales service. Using a Stackelberg game structure, we examine how wholesale pricing, commission rates, and service quality decisions affect the profits of both the manufacturer and 4S dealerships under four distinct combinations (A1–B2). The model incorporates consumer utility sensitivity to both price and service quality and derives optimal strategies under different levels of product substitutability and service cost structures. Results reveal the trade-offs between pricing control and service delegation and offer managerial implications for automotive firms seeking to balance brand consistency, cost efficiency, and customer satisfaction.

Keywords:

Automotive Distribution, Reselling vs. Agency, After-Sales Service, Service Quality, Stackelberg Game

2 - A Comparative Analysis of Display and Shopping Ads on Social Media Platforms

Jiayu Fan, Temple University, Philadelphia, PA, United States, Subodha Kumar

Social media platforms are increasingly adopting monetization strategies that integrate e-commerce functionalities. Platforms like TikTok and Instagram have pioneered the use of shopping ads, which can generate both ad revenue and commission revenue. For started social media platforms, it is unclear for them whether shopping ads are more profitable than display ads. Furthermore, the ad pricing strategy and ad amounts can be different for display ads and shopping ads, which may be affected by the commission rate. However, there is no existing study that can answer these questions. Hence, our study fills this gap by providing a framework to compare the profit, ad price, and ad amount for display ads and shopping ads. Our findings suggest that shopping ads may generate higher profit than display ads even if users dislike shopping ads more than display ads and the product value sold through shopping ads is relatively low. In addition, if users dislike display ads more than shopping ads: 1) the platform may still price display ads at a lower rate compared to shopping ads when the market size of display ads is larger than that of shopping ads; 2) the platform may put more display ads than shopping ads, interestingly, this may happen when users' aversion to display ads is greater than a given threshold meanwhile product value is relatively high. This understanding provides actionable insights for platforms looking to optimize their advertising revenue and commission revenue.

3 - Dynamic Budget Allocation and Sponsored Advertising Policy in E-Commerce with Lost Clicks

Yunzhi Cao, City University of Hong Kong, Hong Kong, Hong Kong, Houmin Yan

Sponsored product is an advertising service provided by a platform for retailers to promote products in E-commerce. With a finite budget of a retailer, this paper studies a joint budget allocation and advertising policy over multiple periods. At each period, the retailer needs to allocate a budget to a financial account for the expense of advertising service during this period. A cost-per-click occurs after setting the bidding price for the expression on the sponsored list, which brings lost clicks if there is no remaining budget in the financial account. We formulate a Markov decision process and establish a dynamic programming reformulation. Incorporating the definition of stochastically concave in midpoint, we provide general and easy-to-verify conditions for a tractable optimal policy, that is, a base allocation listing bidding policy, and characterize its structural properties.

4 - Marketing Mix Modeling as a Strategic Tool: Driving Business Growth

Micky Mukesh Chhabda, University of Connecticut School of Business, Auburn Hills, MI, United States

Marketing Mix Modeling (MMM) offers a powerful analytical framework that quantifies the influence of diverse marketing channels on sales and revenue, empowering organizations to optimize budgets and maximize return on investment. This keynote delves into advanced MMM methodologies that extend beyond traditional reporting, illustrating how innovative modeling approaches can unlock strategic insights and guide transformational marketing decisions.

Attendees will explore the design and execution of MMM projects aligned with business objectives, focusing on interpreting results to make data-driven budget allocations that enhance competitive positioning. The presentation features real-world, cross-industry case studies demonstrating significant improvements in marketing efficiency and measurable business growth driven by MMM insights.

By sharing practical techniques and evidence-based frameworks, this session equips analytics professionals and marketing leaders to champion MMM within their organizations, foster stakeholder trust, and drive impactful, sustainable growth through data storytelling and analytics innovation.

Key Takeaways:

- Understanding the fundamentals of Marketing Mix Modeling
- Designing MMM studies that align with strategic business goals
- Translating complex MMM results into actionable marketing decisions
- Integrating MMM with broader analytics frameworks for comprehensive insights
- Real-world examples demonstrating measurable ROI and business impact

SC48

Building B Level 3 B308

Resilient Decision Making

Invited Session

Decision Analysis Society

Chair: Laixi Shi, California Institute of Technology (Caltech), Pasadena, CA, United States

1 - Nonparametric Policy Improvement in Continuous Action Spaces via Expert Demonstrations**Enrique Mallada, Johns Hopkins University, Baltimore, MD, United States, Agustin Castellano, Sohrab Rezaei, Jared Markowitz**

The policy improvement theorem is a cornerstone of classical reinforcement learning in discrete action spaces, providing critical theoretical guarantees for policy optimization. However, extending these guarantees to continuous action spaces—especially under function approximation—has long been an open challenge, often forcing policy optimization algorithms to rely on conservative updates that slow convergence. In this talk, I introduce a novel nonparametric policy framework tailored for deterministic Markov Decision Processes (MDPs), which admits a provable policy improvement theorem without requiring parametric assumptions. Leveraging mild regularity conditions on the optimal policy, we demonstrate how data from expert demonstrations can be used to construct a nonparametric lower bound on policy value, enabling robust evaluation. This lower bound directly yields a simple yet effective improvement mechanism: add more demonstrations. Moreover, we provide a method to identify regions in the state space where additional data is most needed to meet performance targets. Building on these insights, we propose a policy optimization algorithm that guarantees monotonic improvement of the value lower bound and offers high-probability performance guarantees. This work lays the foundation for a rigorous, data-driven approach to policy improvement in continuous action spaces.

2 - Enhancing Safety in Sequential Decision-Making**Shangding Gu, UC Berkeley, Berkeley, CA, United States**

As artificial intelligence continues to evolve, reinforcement learning has emerged as a powerful tool for training agents to make decisions in complex environments. However, ensuring safety during the learning and deployment phases remains a critical challenge, especially in high-stakes domains like autonomous driving and robotics. This talk will introduce the fundamental concepts behind safe reinforcement learning, discuss why safety is vital in real-world applications, and highlight key methodologies aimed at ensuring safety during the learning process.

3 - Room for Improvement: Optimal and Fair Housing Allocation via Weakly Coupled Markov Decision Processes**Aras Selvi, Imperial College Business School, London, United Kingdom, Bill Tang, Cagil Kocyigit, Phebe Vayanos, Wolfram Wiesemann**

TBA

4 - Hybrid Transfer Reinforcement Learning: Provable Sample Efficiency from Shifted-Dynamics Data**Chengrui Qu, California Institute of Technology, Pasadena, CA, United States, Laixi Shi, Kishan Panaganti, Pengcheng You, Adam Wierman**

Online reinforcement learning (RL) typically requires online interaction data to learn a policy for a target task, but collecting such data can be high-stakes. This prompts interest in leveraging historical data to improve sample efficiency. The historical data may come from outdated or related source environments with different dynamics. It remains unclear how to effectively use such data in the target task to provably enhance learning and sample efficiency. To address this, we propose a hybrid transfer RL (HTRL) setting, where an agent learns in a target environment while accessing offline data from a source environment with shifted dynamics. We show that -- without information on the dynamics shift -- general shifted-dynamics data, even with subtle shifts, does not reduce sample complexity in the target environment. However, focusing on HTRL with prior information on the degree of the dynamics shift, we design HySRL, a transfer algorithm that outperforms pure online RL with problem-dependent sample complexity guarantees. Finally, our experimental results demonstrate that HySRL surpasses the state-of-the-art pure online RL baseline.

SC49

Building B Level 3 B309

Platform and Supply Chain Management

Invited Session

MSOM: Supply Chain

Chair: Rachel Chen, University of California at Davis, 3208 Gallagher Hall, Davis, CA, 95616, United States

Co-Chair: Xuying Zhao, Texas A&M University, College Station, TX, United States

1 - Who Directs the Control Tower? Customer vs. Supplier-Managed Inventory in Supply Chains**Shi Chen, University of Washington Foster School of Business, Seattle, WA, United States, Morris Cohen, Hau Lee**

In recent years, information distortion and incentive misalignment have contributed to suboptimal supply chain performance. As a result, many companies have adopted the "control tower" approach, in which one party assumes responsibility for replenishment decisions across the supply chain. But who should direct the control tower? In some cases, this role is assigned to the customer -- Customer-Managed Inventory (CMI) -- and in others, to the supplier -- Supplier-Managed Inventory (SMI). This paper investigates when it is preferable to adopt one approach over the other. We consider a two-stage supply chain with either a one-to-one inventory structure (a customer with a single warehouse) or a one-to-many structure (a customer with multiple warehouses), and derive the optimal inventory policies and cost-sharing contract parameters under both CMI and SMI. Our analysis reveals that two primary factors drive the choice between CMI and SMI: the shortage penalty cost ratio between the customer and supplier, and the number of customer warehouses. A higher customer-to-supplier penalty cost ratio or fewer customer warehouses tends to favor CMI, while the opposite conditions favor SMI. Secondary factors, including holding costs, lead times, the distribution of demand across warehouses, and demand correlation, also play a role. We illustrate our findings with four industry cases: Barilla, Saturn, Boeing, and HP. The results offer practical guidance for selecting and implementing the appropriate control tower strategy, enabling firms to improve supply chain performance through more effective allocation of decision rights and cost-sharing mechanisms.

2 - Mainstreaming vs. Fairwashing in Fair Trade Operations**Daehoon Jung, University of Florida, Gainesville, FL, United States, Seung-Jae Park**

This study explores the effects of competition on fair trade operations. Specifically, as a benchmark, we first consider a model with single fair trade organization (FTO), i.e., a monopoly FTO. We then extend the analysis to a setting with two competing FTOs, i.e., duopoly FTOs. Although both FTOs share the common goal of alleviating farmers' poverty, they adopt contrasting strategies: one FTO seeks to lower the barriers to fair trade in an effort to mainstream it, while the other maintains stricter

standards to prevent fairwashing (or greenwashing). Following industry practices, we implement two modes of competition and compare them with the monopoly benchmark. We find that competition between FTOs increases the aggregate demand for fair trade certified products (mainstreaming) but results in less aggregate premium (fairwashing). Furthermore, we consider the potential application of blockchain technology in fair trade operations, demonstrating that blockchain offers a promising solution to mitigate the fairwashing effects from competition between FTOs.

3 - Inventory Holdback for Humanitarian Logistics: Reducing Wastage Under Uncertain Demand

Nesim Erkip, Bilkent University, Ankara, Turkey, Vera Tilson

Humanitarian organizations operate under strict donor requirements and financial scrutiny, often facing challenges in managing medical supply distribution due to demand uncertainty and logistical constraints. Expired or excess supplies can strain relationships with donors and host governments, making prepositioning stock without a clear usage plan risky. Reverse logistics for retrieving unused inventory is costly and complex. Motivated by these challenges, we model a humanitarian distribution system as a two-echelon, one-warehouse multiple-retailer (OWMR) supply chain, where demand at retailers is highly variable and not necessarily stationary. Given the potential for wastage, there is an incentive to hold more inventory at a central warehouse rather than at the retailers. We introduce a novel holdback distribution policy that withholds a portion of stock using a non-myopic two-period planning horizon to minimize total costs associated with stockouts and wastage. Computational experiments show that our policy outperforms conventional myopic distribution strategies by effectively reducing stockouts and excess inventory, making it particularly well-suited for volatile demand environments.

4 - Enhancing Sustainability: Smallholder Farmers' Incentives for Sharing Yield Data Information

Jae-Hyuck Park, New Jersey Institute of Technology, Newark, NJ, United States, Hyo Duk Shin

Recent technological advancements have expanded agriculture-related data accessible to small farmers. We examine when smallholder farmers are incentivized to share yield uncertainty information and its sustainability implications. Farmers' decisions are influenced by market structure and information accuracy, highlighting the need for alignment and economic incentives to ensure sustainable land use.

5 - Fulfillment-by-Platform as Strategy: Competitive and Antitrust Implications

Yao Li, Kellogg, Northwestern University, Evanston, IL, United States, Lauren Lu, Wenqiang Xiao, Jan A. Van Mieghem

By tying access to Prime customers and enhanced platform visibility to the use of Fulfillment by Amazon (FBA), Amazon effectively compels third-party sellers to adopt its proprietary logistics service. While this practice may improve operational consistency, it raises significant antitrust concerns by limiting sellers' fulfillment options, thereby reducing cross-platform competition and potentially harming consumer welfare. In this study, we develop a stylized analytical model that incorporates both Amazon's FBA requirement and the broader dynamics of cross-platform competition. We examine settings with one or multiple sellers operating across competing platforms and characterize the conditions under which Amazon's tying strategy may reduce—or, counterintuitively, enhance—consumer welfare. Our findings contribute to the ongoing debate on digital platform governance and the trade-offs between operational efficiency and market fairness.

SC50

Building B Level 3 B310

Recent Advances in Data-Driven and AI-Guided Decision-Making

Invited Session

MSOM: Service Operations

Chair: Hamsa Bastani, Wharton School, Philadelphia, PA, United States

Co-Chair: Stefanos Poulidis, INSEAD, Fontainebleau, France

1 - Deepstock: Reinforcement Learning with Policy Regularizations for Inventory Management

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

In the age of big data and compute, Deep Reinforcement Learning (DRL) has emerged as a promising methodology for inventory management. However, off-the-shelf implementations have shown mixed success. We demonstrate that combining DRL with policy regularizations, grounded in classical inventory concepts such as the “basestock” can significantly improve its effectiveness. In particular, our policy regularizations accelerate training, reduce sensitivity to hyperparameters, and improve overall performance across several DRL algorithms. We also revisit classical DRL policies from the perspective of data usage, comparing those that update based on state-action pairs with those that update based on entire trajectories. We evaluate their efficiency and performance under varying data quality, with a particular focus on the quality of endogenous features. We implement our regularized DRL approach at scale within Alibaba and conduct additional experiments on synthetic data to better understand the benefits of different combinations of DRL policies and regularizations.

2 - Self-Regulated use of AI Hinders Long-Term Learning

Stefanos Poulidis, INSEAD, Fontainebleau, France, Hamsa Bastani, Osbert Bastani

As generative AI tools become widely adopted, educational environments are shifting from \emph{system-regulated} (or paternalistic) to \emph{self-regulated} learning, where learners control when and how to engage with AI tutors. While greater learner agency is often presumed to enhance motivation and learning, access to AI without guidance or guardrails can backfire. We study the impact of self-regulated use of generative AI on long-term learning in a 100-day field experiment with students recruited from chess academies, comparing system-regulated AI assistance with self-regulated use, in which students decide when and how to request help. Students in both conditions improved, but those in the system-regulated condition exhibited significantly greater engagement and learning. We find that highly motivated students self-regulate more effectively and achieve substantially greater skill improvement. In contrast, skill level does not moderate the effect: even skilled students tend to offload cognitive effort and are harmed by self-regulated use of generative AI tools. Our findings highlight how design choices in generative AI deployments can shape long-term learning outcomes and inform the scaffolding of effective AI-mediated learning environments.

3 - Data-Driven Decision Making in Dynamic Systems

Stefan Wager, Stanford GSB, Stanford, CA, United States

This talk will survey recent advances in causal inference and optimal targeting in settings where problem dynamics are of central interest.

SC51

Building B Level 3 B311

Emerging Topics in Socially Responsible Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Somya Singhvi, USC Marshall School of Business, Los Angeles, CA, United States

1 - Disability-Inclusive Workforce Planning

Chen Hu, Xi'an Jiaotong-Liverpool University, Suzhou, China, People's Republic of, Anyan Qi, Yongbo Xiao

Hiring people with disabilities (PWDs) is widely recognized as an essential component of corporate social responsibility (CSR). However, such employment often entails challenges in lower service level or increased training costs due to the generally limited educational attainment of PWDs. This paper studies the hiring and training decisions for a socially responsible firm and examines how these decisions are shaped by government intervention policies.

2 - Sustainable Sourcing of Agricultural Products: Fixed vs. Flexible Premiums

Can Zhang, Duke University, Durham, NC, United States, Vishal Agrawal

Sustainability certifications have gained increasing popularity for agricultural product sourcing. To help improve smallholder farmers' livelihoods, these certifications set a premium that firms must pay to farmers for certified crops on top of the market price. This premium can be structured in different ways: A popular approach adopted by several major certifications is called a fixed premium, under which farmers receive a fixed level of premium independent of the market price. However, this approach has been criticized for not sufficiently protecting farmers when the market price is low. Accordingly, an alternative approach called a flexible premium, under which farmers receive a higher premium when the market price is lower, has been promoted by farmer advocacy groups and recently implemented by some certifications. We analyze and compare the effectiveness of these premium approaches and offer several managerial insights. First, contrary to expectation, we show that the flexible premium approach can lead to a lower expected farmer income than the fixed premium approach. Even when a flexible premium improves farmer income, it can lead to a lower expected sustainable sourcing quantity and firm profit. Second, we nevertheless identify that a flexible premium can lead to a "win-win-win" outcome for farmer income, sustainable sourcing quantity, and firm profit, but only in settings with ample certified supply relative to demand. Finally, motivated by the trend of firms creating their own sustainability labels, we show that such self-labeling can lead to a higher sustainable sourcing quantity but lower farmer income compared to NGO labeling.

3 - Optimizing Inventory for Impact: Large-Scale Inventory Routing to Distribute Clean Cooking Devices in Kenya

Simon Thomä, RWTH Aachen, Aachen, Germany, Sergio Camelo, Dan Iancu, Maximilian Schiffer

Access to clean cooking technologies remains an issue in developing economies. Around 2.1 billion people around the globe rely on polluting fuels for cooking, leading to more than 3 million premature deaths annually. In Kenya, where most households rely on polluting fuels, bio-ethanol stove suppliers use carbon credits to subsidize stove prices. Still, affordability remains a challenge due to high distribution costs caused by highly stochastic and heterogeneous demand distributed over several thousand small retailers in Nairobi alone. We address this challenge by formulating a large-scale Inventory Routing Problem with stochastic demand and propose a new two-step approach to solve it. First, we relax the problem by taking long-term averages to compute target policies and valid lower bounds. Building on these target policies, we derive dynamic routing and replenishment policies that scale well to real-world instances with several thousand stores. Our numerical experiments show that the proposed methods significantly improve over the status quo, leading to significant cost savings.

4 - Optimizing Nonprofit Operations Under Display-Dependent Demand

Yanchong Zheng, Massachusetts Institute of Technology, Cambridge, MA, United States, Yuan Shi, Mahyar Eftekhari

We study a nonprofit warehouse inventory problem motivated by data from a regional U.S. food bank. Many food banks use a shopping-like allocation model where partner agencies select food items directly from warehouse shelves, facing the dual goal of achieving equitable access over time (fairness) and maximizing accepted donations within limited capacity (efficiency). We develop a dynamic inventory model that accounts for display-dependent stochastic demand, heterogeneous resource types, and capacity constraints. We characterize optimal acceptance and display policies that balance fairness and efficiency and identify key drivers of the fairness-efficiency trade-off, including demand uncertainty and resource heterogeneity. Our findings support the design of a decision-support tool now under development with our partner food bank. Backtesting on inventory data shows an over 10% improvement in both fairness and efficiency compared to current practices.

SC52

Building B Level 3 B312

Sustainability in Retail Operations

Invited Session

MSOM: Sustainable Operations

Chair: Fei Gao, Indiana University Bloomington, Bloomington, IN, United States

Co-Chair: Yunlong Peng, University of Warwick- Warwick Business School, Coventry, United Kingdom

1 - Does Second-Hand Clothing Perpetuate Fashion Overproduction?

Xiaoyang Long, University of Wisconsin-Madison, Madison, WI, United States, Luyi Gui, Bo Pang

In this work, we analytically study the impact of a resale platform on fashion consumers' purchase and resale decisions. We analyze the conditions under which the provision of a resale option could inadvertently increase total consumption and production, leading to worse environmental outcomes.

2 - A Novel Return Policy: Negotiating Refunds Without Returns

Tingliang Huang, Haslam College of Business, University of Tennessee, Knoxville, TN, United States, Ziqin Gao, Ganquan Huang, Yan Liu, Yugang Yu

Returns of low-salvage-value products intensify the fundamental return dilemma in retailing industries. Recently, some retailers tried to address this dilemma by adopting the Negotiating Refunds Without Returns (NROR) policy, where consumers keep mismatched products with a negotiated refund without physical returns. To assess the profitability and social implications of NROR, we develop a game-theoretic model based on the Nash bargaining framework.

3 - Reuse and Repeat: the Design of Deposit Return Scheme

Yunlong Peng, University of Warwick- Warwick Business School, Coventry, United Kingdom, Fei Gao, Jian Chen

A growing number of brands are piloting reusable packaging programs managed through deposit-return schemes, in which consumers pay a deposit to borrow reusable packaging and must return it within a specified window to receive a refund. In this paper, we study the optimal design of such schemes and evaluate their economic and environmental implications.

SC53

Building B Level 3 B313

Practice-driven Revenue Management

Invited Session

Revenue Management

Chair: Xiao Lei, University of Hong Kong, Hong Kong, N/A

1 - On Size Substitution and Its Role in Assortment and Inventory Planning

Yi-Chun Akchen, University College London, London, United Kingdom, Felipe Caro

How should (apparel) retailers manage product sizes? For example, if most customers wearing a given shoe size, such as 9.5, are willing to accept a half-size up or down, is it necessary for a retailer to carry that size at all? Additionally, while identical products in different sizes are treated as distinct SKUs in inventory management, they are often aggregated for assortment and strategic planning, although there is no theoretical justification showing that such aggregation approach is sensible. In this paper, we address these fundamental questions about size management, which have remained largely unexplored in the operations literature. We propose a choice model where each customer forms a consideration set based on the in-stock availability of products of her best-fit size and adjacent sizes. Using a real-world dataset from a large footwear retailer, we show that nearly 25% of the unmet demand caused by stockouts spills over to adjacent sizes. We further solve the assortment and inventory optimization problems under the proposed choice model. Our findings show that the optimal assortment remains unchanged regardless of the likelihood that customers might purchase adjacent sizes. We utilize this finding and further show that inventory policies that ignore size substitution can be (asymptotically) optimal when the demand rate is high or the selling horizon is long. We also propose a mixed-integer program to find inventory levels that account for size substitution and obtain higher profits in the low-demand settings.

2 - When Art Auctions Overlap: Revenue Effects of Concurrency in Online Markets

Hester Zhang, IESE Business School, Barcelona, Spain, Abhishek Deshmane

This paper studies the revenue implications of concurrent auctions on an online art auction platform. The shift to online auctions has dramatically reduced transaction costs for bidders, enabling new formats such as the simultaneous auctioning of similar items—what we call concurrent auctions—a feature rarely feasible in traditional in-person settings. While concurrency may reduce auction revenues due to competition among similar items (“cannibalization”), it may also enhance them by improving matching through lower search costs and facilitating portfolio-based bidding by collectors. Using a novel dataset of bid-level information on over 350,000 items from a European online platform that partners with small auction houses, we examine whether concurrency raises auction revenues. Our preliminary analyses find systematically higher revenues for concurrent auctions of artworks by the same artist. Moreover, we observe that revenues initially rise with the number of concurrent auctions, then decline—suggesting that gains from improved matching initially dominate, but are eventually offset by cannibalization effects. We also explore a potential mechanism behind the revenue gains: bidders appear to cross-reference bid increments across similar concurrent auctions, potentially intensifying competition. These findings indicate that online platforms can strategically leverage concurrency to enhance auction outcomes, with broader implications for the design of digital marketplaces.

3 - Dynamic Segmentation and Personalization in in-App Advertising Games

Xiao Lei, University of Hong Kong, Hong Kong, Hong Kong, Jiacheng Chang, Zhixi Wan

In-app advertising (IAA) games represent a unique segment of mobile gaming, where monetization is achieved solely through rewarded ads that players can opt into for in-game advantages. The design and management of these games are intricate due to the high diversity and rapid attrition rates of players. This study introduces a novel framework that dynamically segments users and personalizes game difficulty with the objective of maximizing each user's lifetime value. We detail a method for parameter estimation that accounts for the exogenous shocks frequently encountered in gaming environments. Additionally, we outline an actionable policy derived from Partially Observable Markov Decision Processes (POMDP), designed for straightforward implementation. The efficacy of our approach is demonstrated through a field experiment conducted with a commercial IAA game, where we achieved a 20% increase in revenue when compared to traditional methods employing uniform and fixed difficulty settings.

4 - Robust Ancillary Pricing for Airlines: When a Single Mark-up Outperforms Personalisation

Xiao Jin, National University of Singapore, Singapore, Singapore, Junjie Ren, Chung Piau Teo

In low-cost carriers, ancillary products, such as checked bags and seat selection, are priced with sparse transaction data. We investigate the revenue effects of **assuming homogeneous willingness-to-pay (WTP) noise when, in reality, the noise varies across customers**. Within this misspecified framework we develop and compare two practical rules: (i) a personalised and (ii) a constant feature-based mark-up. Analytical error bounds show that revenue loss grows with the square of the markup error. Numerical experiments and field tests indicate that when the noise model is inaccurate, the constant markup delivers higher median revenue and lower variability than the personalised rule. The findings suggest that, under data sparsity, it may be better off applying a single, conservatively chosen premium rather than pursuing fine-grained personalisation before the WTP distribution is properly identified.

SC54

Building B Level 2 B216

Advanced in Revenue Management and Marketplace Optimization

Invited Session

Revenue Management

Chair: Dragos Florin Ciocan, INSEAD, Fontainebleau, France

1 - Airline Revenue Management with Carbon

Dragos Florin Ciocan, INSEAD, Fontainebleau, France

In order to contain the airline industry's massive environmental footprint, regulations impose that airlines who exceed their emission thresholds pay into emissions trading systems. This effectively imposes a carbon cost over the flights an airline operates. We study how the presence of these costs affects pricing policies and network layouts for airlines.

2 - A Distributed Supply Chain: Inventory and Capacity Analysis

Victor Araman, American University of Beirut, Beirut, Lebanon, Sasha Ghaya

We consider a simple supply chain model consisting of a retailer facing a constant demand from end-consumers while being supplied by a distributed network of suppliers (DNS) made of local, heterogeneous artisans, each operating as independent entities. Within this setting, the retailer determines the adequate base-stock level, while each supplier decides on the amount of capacity to allocate to the retailer. We specifically model the supply function as an infinite server queue. We suggest an approximation of the problem's formulation obtained through a heavy traffic limiting regime, to which we apply a static mean field game's approach. We analyze first the centralized system as a benchmark and obtain a closed form solution to the decentralized problem. As we do so, we quantify the inefficiency of the decentralized setting and suggest a simple linear transfer payment-type contract that achieves the first best solution.

3 - Improving Housing Market Outcomes via Auditing and Financial Facilitation: a Platform-Based Model of Ngo Intervention

Irmak Dai, INSEAD, Fontainebleau, France, Atalay Atasu, Dragos Florin Ciocan, Antoine Desir

Housing markets in low-income settings often suffer from poor-quality construction, informal housing practices, and information asymmetry between builders and beneficiaries. This leads to inefficient allocation, where even high-budget beneficiaries may receive low-quality housing. We study the role of a nonprofit intermediary that aims to improve market outcomes through two costly interventions: (i) auditing builders to deter misrepresentation, and (ii) facilitating beneficiary access to external financial resources. We develop a stylized model in which builder qualities and beneficiary budgets are continuously distributed, and the intermediary operates under imperfect observability. Preliminary results indicate that no single intervention is sufficient to improve outcomes across the market. Instead, coordinated action—engaging NGOs, public authorities, and financial institutions—appears necessary to address the underlying inefficiencies. This work contributes to the literature on allocation under asymmetric information and platform-enabled interventions in under-regulated markets.

4 - Market Design of Contest Platforms

ZONGSEN YANG, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Xiao Lei, Pin GAO, Zizhuo Wang

We study the market design problem in an online contest platform (e.g., GGPoker) that repeatedly organizes large-scale contests to optimize long-term player retention. We build a novel dynamic model where the players' active probability depends on their skill levels and engagement states. Active players decide whether to participate based on their gains in each round. The platform must incentivize user participation and match them to maximize player engagement. When players' incentives are disregarded, we propose an outcome-based matching scheme that reallocates the number of winning points among players by matching high-level players with low-level ones. This scheme departs from traditional room-based approaches and can be framed as a downstream network flow problem. We prove that the common skill-based matching is optimal when games are highly competitive/non-competitive, or engagement states are limited. However, it could perform poorly compared to the optimal cross-level matching. We also find that when players suddenly disengage after loss streaks, the optimal policy would transfer winning points from the most and least engaged players to the second-least engaged players. However, the optimal policy may completely sacrifice the interests of low-level players and discourage them from joining in the contests. To resolve this, we design a token-based mechanism to incentivize players and recover the first-best outcome. Our results remain valid in the contexts of limited transfer or online arrivals. Our analysis provides actionable insights and practical solutions for contest platforms seeking to better entertain players.

SC55

Building B Level 3 B314

Platforms and Marketplaces

Invited Session

Revenue Management

Chair: Ilan Morgenstern, Yale University, New Haven, United States

Co-Chair: Zhen Lian, Yale School of Management, New Haven, CT, United States

1 - Tagging the Shift: Effect of Large Language Models on Knowledge Communities

Neha Sharma, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States, Simin Li

The widespread adoption of Large Language Models (LLMs) is profoundly altering human-AI interaction, raising critical questions about their impact on online collaborative knowledge ecosystems. We empirically investigate these dynamics using a comprehensive dataset from a large online community for programming and technical expertise. Employing a difference-in-differences framework and network analysis of tag co-occurrence, our study explores how LLMs influence user behavior and the structural organization of knowledge within such platforms. We aim to understand the evolving nature of collective intelligence and human-AI co-evolution in these digital spaces.

2 - The Price of Prestige: Exclusive Pricing and Selling of Luxury Bands and Resale Market Dynamics

Zhe Liu, Imperial College London, London, United Kingdom, Anran Li, Rachel Zhang

Luxury brands like Hermès and Chanel balance exclusivity with revenue through distinct strategies. Hermès requires customers to purchase ancillary products before accessing premium items like Birkin bags, while Chanel sets higher direct prices without bundling requirements. This study examines how luxury brands should design pricing mechanisms considering resale market dynamics. We develop a novel model capturing customers with dual valuations for premium and ancillary products, creating three segments: primary buyers, resellers, and resale market buyers. Our game-theoretic analysis shows how optimal pricing emerges from equilibrium interactions between these customer types and reveals why seemingly restrictive strategies can be optimal. We provide insights for luxury brands to enhance prestige, revenue, and customer loyalty while navigating primary and secondary market dynamics.

3 - Who Ate the Lunch? A Belief Propagation Algorithm for Identifying Refund Hunters

Chenkai Yu, Columbia University, New York, NY, United States, Arpit Agarwal, Hongyao Ma

For online platforms such as UberEats and DoorDash, a central challenge in customer service is the lack of ground truth: when a customer reports that an order was never received, it is difficult for support agents to determine if the driver kept the food, a passer-by took the delivery, or the customer is falsely claiming a missing order. This fundamental uncertainty often results in platforms shouldering refunds and appeasement costs without holding either side of the market accountable. In this work, we propose a belief propagation (BP) algorithm for identifying strategic customers and drivers, considering both their frequency of platform use and the trustworthiness of the participants they had interacted with. In large markets, we show that our proposed BP algorithm approximately recovers the correct types on the driver side, and achieves the highest statistical power on the customer side. Additionally, there is an exponentially large gap in the ratio of false positive rates between our algorithm and a naive benchmark that only relies only on the average past refund rate of each participant. Extensive experiments on both synthetic data and data from our industry collaborator — a major Southeast Asian platform — demonstrate that the BP algorithm provides substantial and robust accuracy improvements.

4 - Learning Competition Status with Small Samples in Gig Economy Platforms

Zhen Lian, Yale School of Management, New Haven, CT, United States

Gig economy platforms such as Uber and Lyft operate in dynamic and competitive markets where drivers may receive offers from multiple platforms. A central challenge is determining optimal compensation when drivers' reservation wages are both heterogeneous and influenced by competitors' actions—factors typically hidden from the platform. For example, temporary pay incentives offered by Uber can raise drivers' willingness-to-accept on Lyft, yet Lyft has no direct visibility into such promotions. We study how a ride-hailing platform can make adaptive pay decisions using small samples and real-time driver responses. We model the platform's problem as a sequential decision process in which pay offers are adjusted based on observed acceptances and rejections, with rejections serving as implicit signals of external competition. We analyze three settings— independent, fully correlated, and mixed drivers—and characterize the structure of optimal pay policies using dynamic programming. Our results highlight that optimal wage sequences are non-decreasing across drivers, and we quantify the trade-off between flexibility in wage setting and the value of information about competition. We further show how a platform can infer competitor activity from limited data and adaptively refine its pay strategy. This framework provides structural insights into wage design and offers a practical approach for platforms to optimize driver pay in uncertain, competitive environments.

SC56

Building B Level 3 B315

Modern AI and Marketplaces

Invited Session

Auctions and Market Design

Chair: Yeganeh Alimohammadi, USC, Los Angeles, CA, United States

1 - Optimal Selection Using Algorithmic Rankings with Side Information

Kathleen Donahue, MIT, Cambridge, MA, United States, Nicole Immorlica

Motivated by online platforms such as freelance job markets, we study an agent choosing from a list of candidates, each with a hidden quality that determines match value. The agent observes only a noisy ranking of the candidates plus a binary signal that indicates whether each candidate is "free" or "busy." Being busy is positively correlated with higher quality, but can also reduce value due to decreased availability. We study the agent's optimal selection problem in the presence of ranking noise and free-busy signals and ask how the accuracy of the ranking tool impacts outcomes. In a setting with one high-valued candidate and an arbitrary number of low-valued candidates, we show that increased accuracy can result in reduced social welfare. This can occur for two reasons: agents may be more likely to make offers to busy candidates, and (paradoxically) may be more likely to select lower-ranked candidates when rankings are more indicative of quality. We further discuss conditions under which these results extend to more general settings.

2 - Data Informativeness in Decision-Making

Amine Bennouna, Kellogg, Northwestern University, Cambridge, MA, United States, Omar Bennouna, Saurabh Amin, Asuman Ozdaglar

We study the fundamental informational value of a dataset with respect to a given decision-making task under uncertainty. Given a task and an uncertainty set, we characterize which datasets are sufficient to recover the task's optimal solution. These results enable us to introduce a novel principled approach for the important problem of task-specific data collection.

3 - Breaking Algorithmic Collusion via Simple Defections

Meena Jagadeesan, UC Berkeley, Berkeley, CA, United States, Natalie Collina, Eshwar Ram Arunachaleswaran

When firms deploy algorithms to automate pricing decisions, the implicit competition between these algorithms can lead to market prices that are higher than in a competitive market. In this work, we investigate whether high prices persist even when some firms defect to simpler heuristics. Our main finding is that even a single firm defecting to a simple heuristic strategy can significantly drive down the market prices, though not necessarily all the way down to the competitive price. We specifically

analyze a repeated Bertrand pricing game under defections to no-regret strategies and FollowTheLeader (FTL). We show that when the lowest-profit firm defects to a no-regret strategy, the worst-case market price scales approximately inversely with the number of other firms in the market. Furthermore, we extend our results to defections by multiple firms at once, and we also show that FTL can lead to strictly higher prices than no-regret defections. Taken together, our results illustrate how algorithmic collusion can break down when even a small number of firms do not adopt pricing algorithms, as long as there are sufficiently many other firms in the market.

4 - Balancing Optimality and Practicality: Bidding Strategies for Value Maximizers in Repeated Discriminatory Auctions

Sourav Sahoo, MIT, Cambridge, MA, United States, Negin Golrezaei

We study repeated multi-unit discriminatory (pay-as-bid) auctions, commonly used to sell treasury securities and commodities. Our goal is to design bidding strategies for value-maximizers—those seeking to maximize total value across rounds while satisfying per-round return on investment (RoI) constraints. Motivated by real-world platforms, we assume buyers submit bids in an m -uniform format. We first characterize *safe* bidding strategies—those that always satisfy RoI constraints regardless of competitors' bids—and show they correspond to simple no-overbidding conditions. Despite this simplicity, safe strategies are *lossless*: they achieve the same value as any RoI-feasible strategy. We then develop a polynomial-time online learning algorithm that learns optimal safe strategies with sublinear regret under both full information and bandit feedback. We next analyze a finite subset of safe strategies, called *base strategies*. While the worst-case performance of the best base strategy is only a $1/m$ -fraction of the optimal safe strategy's value, our empirical analysis on Austrian treasury bond auctions reveals that base strategies far outperform this bound in practice. Furthermore, they exhibit faster learning over time and benefit from lower computational and space complexity, making them a practical and efficient alternative to the theoretically optimal safe strategies. Finally, we present a result that bridges these two classes of strategies—any m -uniform safe strategy can be expressed as a convex combination of at most m base strategies.

5 - Improved Approximations for Stationary Bipartite Matching: Beyond Probabilistic Independence

Tristan Pollner, Stanford University, Palo Alto, CA, United States, Amin Saberi, Ali Aouad, Alireza Amaniamedani

We study stationary online bipartite matching, where both types of nodes—offline and online—arrive according to Poisson processes. Offline nodes wait to be matched for some random time, determined by an exponential distribution, while online nodes need to be matched immediately. This model captures scenarios such as deceased organ donation and time-sensitive task assignments, where there is an inflow of patients and workers (offline nodes) with limited patience, while organs and tasks (online nodes) must be assigned upon arrival.

We present an efficient online algorithm that achieves a $(1-1/e+\delta)$ -approximation to the optimal online policy's reward for a constant $\delta>0$, simplifying and improving previous work by Aouad and Saritaç (2022). Our solution combines recent online matching techniques, particularly pivotal sampling, which enables correlated rounding of tighter linear programming approximations, and a greedy-like algorithm. A key technical component is the analysis of a stochastic process that exploits subtle correlations between offline nodes, using renewal theory. A byproduct of our result is an improvement to the best-known competitive ratio—that compares an algorithm's performance to the optimal offline policy—via a $(1-1/\sqrt{e+\eta})$ -competitive algorithm for a universal constant $\eta>0$, advancing the results of Patel and Wajc (2024).

SC57

Building B Level 3 B316

Social Networks and Predictive Insights: Analysis, Markets, and Models

Invited Session

Social Media Analytics

Chair: Tong Wang, Yale University, New Haven, CT, United States

1 - Social Media and Ride Hailing Demand: An Empirical Analysis

Adrija Majumdar, Indian Institute of Management Ahmedabad, Ahmedabad, India, Sandip Chakrabarti

Can content generated by users on social media be leveraged as a strategic tool to adjust service quality and performance to attract and retain customers—especially in industries like shared mobility, where customer experience shapes demand? This study investigates whether the volume and sentiment of social media discussions can help predict ridership demand for app-based ride-hailing services, particularly during volatile periods when traditional forecasting methods fall short. We analyze over fifteen months of tweets about Uber rides posted on the X platform in a major U.S. city during the Covid-19 pandemic. Our findings show that chatter from the previous day positively influences ridership. We further segment the analysis into high- and low-uncertainty phases, revealing that social media discussions have a stronger impact on consumer choices during uncertain times (such as before vaccine announcements), when customers tend to rely more on online information and peer experiences for decision-making. We also identify conversation themes that influence ridership and find that positive discussions boost demand. This research highlights the potential for private ride-hailing providers to use social media data to forecast short-term demand trends, particularly during unpredictable conditions. Public transport agencies can also leverage social media data for shared transportation planning and operations.

2 - The Curse of Rashomon Effects - from Model Explanation to Data Misinterpretation

Ronilo Ragodos, University Of Iowa, Iowa City, IA, United States

Machine learning models have been increasingly used in business research. However, most state-of-the-art machine learning models, such as deep neural networks and XGBoost, are black boxes in nature. Therefore, post hoc explainers that provide explanations for machine learning models by, for example, estimating numerical importance of the input features, have been gaining wide usage. Despite the intended use of post hoc explainers being explaining machine learning models, we found a growing trend in business research where post hoc explanations are used to draw inferences about the data. In this work, we investigate the validity of such use. We investigate with extensive experiments whether the explanations obtained by the two most popular post hoc explainers, SHAP and LIME, provide correct information about the importance of features for different interpretations of feature importance. In particular, we investigate how well SHAP and LIME can tell us features' global and local importance, how useful they are for prediction, and how effective they are in pursuit of actionable recourse. We refer to an explainer's utility in these respects as data-alignment. We then identify what factors influence the alignment of explanations and conclude that it is often not appropriate to infer data insights from post hoc explanations. We articulate appropriate alternative uses, the most important of which is to facilitate the proposition and subsequent empirical investigation of hypotheses. The goal of this paper is to caution business researchers against translating post hoc explanations of machine learning models into potentially false insights and understanding of data.

3 - Predicting Team Success with Representation Learning from Multi-Relational Collaboration Graphs—A Case of Nfl Coaching Teams

Seyong HA, University of Iowa, Iowa City, IA, United States

Despite its potential for insights into predicting team success, coaching collaboration has received little attention. This research explores how multi-relational structures in coaching collaboration graphs influence team success using the National Football League dataset. We propose a graph representation learning framework that captures these complex dynamics and offers a new perspective on coaching collaborations. This framework predicts a team's seasonal success by learning multi-relational collaboration representations and utilizes individual coach features such as historical performance. Our results show that leveraging both relational and individual attributes leads to an enhanced predictive power.

Prior work has typically aggregated coaching collaborations into a single network. In contrast, our framework utilizes five distinct collaboration networks that offer a richer representation. These include embeddings from three separate collaborations (supervision, reporting, and peer) and two more with different weighting schemes to the original collaboration graph. We apply an attention mechanism over the embeddings to formulate context-rich coach representations. Subsequently, the framework synthesizes team representation embeddings by attentively aggregating these individual coach representations. Attention in both aggregation steps allows us to identify and emphasize the most predictive collaboration types and key coaches while enhancing the explainability of influential factors.

Applying this approach to the NFL coach dataset from the 1963 to the 2019 season demonstrates that we can achieve better performance in terms of both explainability and accuracy via the proposed framework. Our results shed light on the vital role of coaching collaboration and reveal how it can be used to predict team success in sports.

4 - The Geometry of Persuasion: Mapping Opinion Dynamics in Embedding Space

Yen-Shao Chen, Yale University, New Haven, CT, United States, Tauhid Zaman

Persuasion—the process of changing opinions or behaviors—exhibits surprisingly elegant geometric patterns when analyzed through the lens of large language models. We investigate a paradigmatic scenario: a salesperson persuading a customer to purchase a product. By embedding conversational messages into high-dimensional space, we discover that successful persuasive interactions trace distinctive curves through this landscape. Different persuasion strategies manifest as unique trajectories, revealing an underlying geometry to the art of influence. We construct a vector field in the embedding space that identifies persuasion paths, offering both theoretical insights into the mechanics of opinion change and practical applications for understanding effective communication strategies. This geometric framework provides a novel perspective on how minds navigate the space of beliefs during persuasive encounters.

SC58

Building B Level 3 B317

Advances in Scheduling and Project Management

Contributed Session

Contributed

Chair: Kelvin Ekun, University of Massachusetts, Amherst, Amherst, United States

1 - Reinforced Stable Matching for Crowd-Sourced Delivery Systems Under Stochastic Driver Acceptance Behavior

Shixuan Hou, Ivey Business School, Western University, London, ON, Canada, Bissan Ghaddar

Crowd-Sourced Delivery Systems (CDS) depend on occasional drivers to deliver parcels directly to online customers. These freelance drivers have the flexibility to accept or reject orders from the platform, leading to a stochastic and often unstable matching process for delivery assignments. This instability results in frequent rematching, delayed deliveries, decreased customer satisfaction, and increased operational costs, all highlighting the critical need for improved matching stability within CDS. While traditional stable matching theory provides a foundation, it primarily addresses static and deterministic scenarios, making it less effective in the dynamic and unpredictable environments typical of CDS. Addressing this gap, this research extends the classic Gale–Shapley (GS) stable matching algorithm by incorporating tailored compensations for drivers, incentivizing them to accept assigned orders and thus improving the stability of matchings, even with the inherent uncertainties of driver acceptance. We prove that the proposed mechanism can generate reinforced stable matching results based on tailored compensation values. Our numerical study shows that this reinforced stable matching approach significantly outperforms traditional methods in terms of both matching stability and cost-effectiveness. It reduces the order rejection rate to as low as 1% and cuts operational costs by up to 18%.

2 - Order Assignment and Two-Stage Integrated Scheduling in Fruit and Vegetable Supply Chains

Gang Wang, University of Massachusetts Dartmouth, North Dartmouth, MA, United States

Fruit and vegetable supply chains require tight integration to maintain freshness during post-harvest, production, and distribution. This paper addresses the scheduling problem in a two-stage flow shop within the supply chain, involving procurement, production, distribution, and order assignment with contract growers, processors, and demand points. The first stage focuses on production and distribution schedules from growers to processors, while the second stage involves assigning demand points to processors and scheduling production and distribution from processors to demand points. We propose a mixed-integer linear programming model for these processes and present two strongly polynomial algorithms to solve the sub-problems. Additionally, we develop a two-phase heuristic algorithm that integrates both stages. Computational experiments show that our algorithms outperformed CPLEX in large-scale instances while also analyzing the effects of waste reduction, processing costs, and shipping rates on supply chain performance.

3 - A Hybrid Genetic Programming for Online Rescheduling of Semiconductor Manufacturing Systems

Yutong Su, Arizona State University, Tempe, AZ, United States, Feng Ju, Nital Patel

Semiconductor manufacturing assembly systems are subject to random disruptions such as machine failures, which can lead to delays and time window violations that render the pre-determined master schedule suboptimal or even infeasible. Recreating a new master schedule is time-consuming and may incur greater losses than the original disruption. To mitigate the negative impact of machine failures, we propose a novel hybrid Genetic Programming approach for short-horizon rescheduling in real-world semiconductor manufacturing systems. The proposed method introduces custom-designed features, referred to as terminals, specifically tailored for handling machine failures, which make the rescheduling process inherently interpretable, in contrast to many meta-heuristics or machine learning methods where the rationale behind adjustments is often difficult to discern. Experimental results demonstrate that the proposed approach significantly reduces delays within a reasonable computation time, thereby improving overall system efficiency.

4 - Creator Incentives in Recommendation Systems: a Cooperative Game-Theoretic Approach to Multi-Agent Bandits

Ramakrishnan Krishnamurthy, New York University, New York, NY, United States, Arpit Agarwal, Lakshmi Subramanian, Max Nickel

In modern online recommender systems, user interactions with content serve as implicit learning signals, enabling platforms to generalize preferences across users and content. Consequently, content creators influence not only their own outcomes but also those of others—user feedback on one creator’s content can affect how the system recommends others’ content, impacting their exposure and revenue across the ecosystem. Recognizing this interdependence, we study how observation sharing (collaboration) among creators (agents) affects learning efficiency and investigate the conditions in which such collaboration is collectively rational and equitable.

To this end, we model the problem as a collaborative variant of the multi-agent stochastic linear bandit framework, where agents individually interact with a common environment but may form coalitions to share observations and improve learning efficiency. We formalize this as a transferable utility (TU) cooperative game, where the value of a coalition corresponds to the negative sum of bandit regrets incurred by its members. This framework allows us to examine how algorithmic design and structural assumptions about agents affect collaboration incentives.

We show that under some algorithmic conditions, the induced TU games exhibit desirable properties: for identical agents, the game is convex and admits a non-empty core containing the Shapley value, ensuring stable and equitable collaboration. For heterogeneous agents, we demonstrate core non-emptiness and propose a simple, implementable payoff mechanism that satisfies all but one of the Shapley value axioms. Experiments on MovieLens-100k-derived bandit problems highlight where empirical payouts align or diverge from theoretical ideals, offering insights for incentive-aligned collaborative learning design.

5 - Online Lead TIME Quotation Under Delayed Feedback: Theoretical Guarantees and Empirical Evaluation with Offline Production Scheduling

Kelvin Ekun, University of Massachusetts, Amherst, Amherst, MA, United States, Ana Muriel

As customer order inquiries arrive over time, make-to-order (MTO) firms must provide competitive lead time quotations on the spot and manage their capacity to fulfil customised orders in a timely fashion under high uncertainty on future demand and when/whether customers will accept the given quotations. The firm must balance between quoting longer lead times that might result in the loss of the order if customers are not willing to wait and quoting shorter lead times that may end in significant delay penalties. We first consider the case of customers arriving and receiving quotations sequentially in a single sales period, while customer final acceptance decisions and production occur over the subsequent N periods. We later extend this model to allow for customer arrivals throughout the production periods. In both settings, we propose an online primal-dual quotation algorithm to maximise the total expected profit from arriving orders in a single-machine environment, showing it achieves a competitive ratio of $1/(1+\beta)$, where β is the dual-update parameter. We analyse the algorithm’s performance with periodic production scheduling against benchmarks given by naive conventional rules and an all-knowing offline oracle. Our extensive computational experiments suggest that the performance of the primal-dual algorithm

is very sensitive to the choice of dual-update parameters. Combining worst-case analysis with empirical testing, we identify robust parameter settings and show significant savings relative to our benchmarks.

SC59

Building B Level 3 B318

User Behavior and Social Interaction

Contributed Session

Contributed

Chair: Mohammad Arif Ul Islam, University of North Carolina Greensboro, Greensboro, NC, 27413, United States

1 - Unintended Consequences of Status Effect in Reputation Systems

Yingxin Lin, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, Jihao Hu, Dapeng Xu, Xiaoquan Zhang

To encourage users to contribute higher-quality reviews, many review platforms implement community status systems (e.g., user levels, badges, ranks) as reputational incentives. A reviewer's community status also serves as an important signal for both consumers and marketers in their decision-making processes. In practice, both marketers and consumers tend to place greater trust in reviewers with a higher status. However, it remains unclear whether this is a prudent strategy. This study aims to uncover the existence of status-related preferences among marketers and consumers and, more importantly, to examine whether they truly benefit from this strategy. Drawing on both observational and experimental data, we demonstrate that, on one hand, marketers are more likely to invite high-status (vs. low-status) reviewers to participate in incentive campaigns; on the other hand, reviews posted by high-status reviewers have a greater influence on consumers, even when the review content is identical. Furthermore, we find that high-status reviewers are more inclined to provide lower ratings, suggesting a downward bias in their online rating behavior. This bias diminishes the overall helpfulness of online reviews for both consumers and marketers. We validate these findings through a series of robustness checks and further explore the mechanisms underlying this bias. Finally, we discuss the theoretical and managerial implications of our findings.

2 - Reddit Users Unleashed - Understanding User Behaviour and Their Impact on Meme Stocks

Inez Maria Zwetsloot, University of Amsterdam, Amsterdam, Netherlands, Simon Trimbom

Online social networks play an important role in financial markets and facilitate information sharing among investors for meme stocks, as was shown vividly in the recent Reddit hype. In the literature either a theoretical model is build to characterize the role of individual social network users on financial markets or aggregated empirical models are tested on small networks, leaving a gap for investigating this relationship empirically for large scale networks like Reddit. We study this relationship empirically with network data from WallStreetBets, a subreddit, via a model we term the Interpretable Social Network and Market Behaviour Model (ISBM); a novel model which captures heterogeneous individual user behaviour and links this through innovative change point modelling to the stock market. The model is scalable and includes both influential nodes as well as followers. We find that 1) both influential users as well as followers impact market behaviour, 2) changes in the network are observed well before the market frenzy, 3) changes in users behaviour can explain the market movements and improves the adj. R2 from modelling the return by 630%. Our results lead to insights and recommendations for regulators on effects of meme stocks on markets.

3 - A Game-Theoretic Analysis of System Design in Multiplayer Video Games with A Social Comparison Effect

Haruto Morimoto, Keio University, Yokohama, Japan, Nobuo Matsubayashi

Multiplayer video games have increasingly received consumers' attention, and many games have been distributed online through platforms such as Steam. However, when heterogeneous consumers interact with each other through multiplayer games on various hardware, social comparisons are likely to occur, and they may experience psychological loss or gain. These benefits and costs of social comparisons are due to heterogeneity in terms of the competitive strength of the hardware they have, which has been examined in the literature, as well as their own individual competitive strength.

This study investigates the strategic interaction between one game content developer (e.g. Tencent) and two competing hardware manufacturers (e.g. Nintendo and Sony) in the presence of such social comparison effects. The manufacturers individually choose to adopt either a closed system, whereby their game content is offered only to consumers who have purchased their hardware, or an open system, whereby all consumers on the platform (e.g. steam) can play the game content. It is assumed that social comparisons among consumers due to differences in their hardware only occur under a closed system. The open system is only sustained in equilibrium when both manufacturers choose it.

We show that the open system benefits the manufacturer that offers a superior hardware, while the closed system benefits the manufacturer that offers an inferior hardware. However, we find that the open system can be sustained in the equilibrium when the degree of the social comparison effect is intermediate, which also maximizes the total profit for the channel parties.

4 - Gg or Rage Quit? the Effect of in-Game Peer Toxicity on Self-Toxicity in Competitive Online Gaming Across Sequential Matches

Sunan Qian, University of Notre Dame, Notre Dame, IN, United States, Corey Angst, Yoonseock Son

With the growing popularity of online video games, toxic in-game communication has become a persistent challenge, threatening player retention, community health, and overall gaming experience. Understanding the dynamics of how this toxicity spreads is crucial for designing effective moderation strategies. Using data from 1,358 Dota 2 players across 51,771 matches (Dec 2024 – Jan 2025), we explore how exposure to toxicity in past matches influences a focal player's self-toxicity in subsequent games. We examine whether this contagion effect changes over time, how it is shaped by the outcome of a match (win/loss), and whether certain types of players are more susceptible to adopting toxic behaviors than others. We use a BERT-base natural language model to measure explicit and implicit textual toxicity. Results from our empirical analysis show that toxicity spreads from both teammates and opponents, with teammates exerting over 10 times the negative effect as opponents. Cooldown time between matches does not stop the spread of explicit toxicity, and longer breaks actually increase the spread of implicit toxicity from teammates, suggesting that implicit toxicity may operate more as a persistent habit than a reactive behavior. Toxicity contagion persists regardless of prior match outcomes, even though the peer effect from opponent explicit toxicity is weaker after a win. High-skill players resist opponents but absorb teammate influence, and active players are more susceptible to opponent peer effects. Our findings encourage gaming companies to adopt tailored interventions to address implicit and explicit toxicity through distinct strategies like habit-breaking nudges and rapid response systems.

5 - Can We Talk?: Artificial Emotional Intelligence Chatbots and Loneliness Reduction

Mohammad Arif Ul Islam, University of North Carolina Greensboro, Greensboro, NC, United States, Nikhil Mehta

Artificial Emotional Intelligence (AEI) has become a game-changing development in digital communication by combining affective computing and natural language processing to mimic human empathy. In contrast to traditional rule-based systems, AEI-driven chatbots now provide dynamic, emotionally intelligent interactions. These changes are particularly relevant to the problem of loneliness, which is a complex emotional experience characterized by a sense of emotional and social loneliness. Since loneliness is strongly associated with negative consequences including anxiety, depression, and cardiovascular problems, it is becoming more widely acknowledged as a public health issue. AEI chatbots present a promising intervention by facilitating emotionally resonant conversations that enhance users' psychological well-being. A central mechanism influencing their effectiveness is perceived emotional authenticity (PEA)—the extent to which users regard the chatbot's emotional responses as

sincere and emotionally aligned. High PEA encourages emotional transparency, user happiness, and trust—all of which are critical for applications related to mental health. In order to establish this authenticity, empathy—both generation and simulation—are essential. This study examines how AEI chatbots' empathy mechanisms help users perceive emotional authenticity and, in turn, how this perception affects reductions in social and emotional loneliness. It also looks at how trust may either increase or decrease the effect of PEA on loneliness outcomes. The work intends to contribute to scalable, user-centered methods for emotional support and loneliness reduction by incorporating knowledge from affective computing, human-computer interaction, and mental health research.

SC60

Building B Level 3 B319

AI Performance and Evaluation

Contributed Session

Contributed

Chair: YITING TANG, University of Sydney, 26 cadigal avenue, Sydney, 2000, Australia

1 - On the Statistical Capacity of Deep Generative Models

Edric Tam, Stanford University, San Francisco, CA, United States, David Dunson

Deep generative models are routinely used in generating samples from complex, high-dimensional distributions. Despite their apparent successes, their statistical properties are not well understood. A common assumption is that with enough training data and sufficiently large neural networks, deep generative model samples will have arbitrarily small errors in sampling from any continuous target distribution. We set up a unifying framework that debunks this belief. We demonstrate that broad classes of deep generative models, including variational autoencoders and generative adversarial networks, are not universal generators under a variety of statistically relevant divergences. Under the predominant case of Gaussian latent variables, these models can only generate concentrated samples that exhibit light tails. Using tools from concentration of measure and convex geometry, we give analogous results for more general log-concave and strongly log-concave latent variable distributions. We extend our results to diffusion models via a reduction argument. We use the Gromov-Levy inequality to give similar guarantees when the latent variables lie on manifolds with positive Ricci curvature. These results shed light on the limited capacity of common deep generative models to handle heavy tails. We illustrate the empirical relevance of our work with simulations and financial data.

2 - Avalon: An Agentic Methodology for Multi-Scale, Quantifiable Verification of Research Reports

Murtaza Nasir, Wichita State University, Wichita, KS, United States

Increasingly, research outputs are being augmented or even completely generated by AI systems, and new methods for robust and scalable quality evaluation are required. While traditional peer review is invaluable, it is fraught with issues of subjectivity, scalability, and generating fine-grained, quantifiable feedback that can be used for iterative improvement or training of generative AI. In this paper, we propose AVALON (Agentic Verification for Logically-grounded Objective N-dimensional assessment), a novel agentic system to automatically verify and quantify the quality of research reports, human- or AI-authored. AVALON takes as input a research report and its cited references (optionally with the original user-prompt) and evaluates it using a multi-agent system at multiple scales, including broad assessments of structure, style, and coherence, as well as granular, claim-by-claim factual verification, where agents attempt to retrieve and analyze text from referenced sources. AVALON further includes agents that can execute code to verify quantitative analyses, and search existing literature to find potentially overlooked research. The methodology results in a multi-dimensional quantitative scorecard and qualitative, peer-review-style comments. The goal of AVALON is to help address an important gap in providing an objective, reproducible, and detailed verification method, to enable standardized quality evaluation and a quantifiable reward signal for training and advancement of state-of-the-art AI research systems.

3 - Toward a Trustworthy Optimization Modeling Agent via Verifiable Synthetic Data Generation

Vinicius Lima, IBM Research, Yorktown Heights, NY, United States, Dzung Phan, Jayant Kalagnanam, Dhaval Patel, Nianjun Zhou

We present a framework for training trustworthy large language model (LLM) agents for optimization modeling via a verifiable synthetic data generation pipeline. Focusing on linear and mixed-integer linear programming (LP/MILP), our approach begins with structured symbolic representations and systematically produces natural language descriptions, mathematical formulations, and solver-executable code in multiple optimization languages. By programmatically constructing each instance with known optimal solutions, the pipeline ensures full verifiability and enables automatic filtering of low-quality demonstrations generated by teacher models, without the need for annotation by optimization experts. Each dataset instance includes a structured representation of the optimization problem, a corresponding natural language description, the verified optimal solution, and step-by-step demonstrations — generated by a more powerful teacher model — that show how to model and solve the problem across multiple optimization modeling languages. This enables supervised fine-tuning of open-source LLMs specifically tailored to optimization tasks. Empirical evaluation shows that models trained using our pipeline achieve state-of-the-art performance on standard optimization modeling benchmarks through majority-vote inference and multi-language cross-validation. Our approach provides a scalable, verifiable, and principled path toward building reliable LLM agents for real-world optimization modeling applications.

4 - Genai and Decision Making - An Eye-Tracking Study

Shehara Grabau, Auburn University, Auburn, AL, United States, Pankush Kalgotra

With the rise of Generative AI (GenAI) in e-commerce, more platforms are integrating GenAI tools such as review summarization and chatbots to assist consumers in decision making. Consumers experience less information overload and greater decision satisfaction when exposed to a manageable amount of user reviews. This suggests that GenAI review summarization can reduce the cognitive burden on consumers. Recent work is beginning to uncover how GenAI summaries and their valence influence decision making. For instance, AI-generated summaries have been shown to positively impact product sales, with their effectiveness shaped by factors such as number of reviews, average rating, and rating dispersion. These findings suggest GenAI tools influence purchasing decisions under varying conditions. However, such tools are often examined in isolation. The combined effect of multiple GenAI-powered tools, as in real-world settings, remains underexplored. Our study investigates how GenAI review summaries and chatbot conversations collectively influence consumer decision making. While review summaries provide a quick overview of public opinion, chatbots offer tailored responses to specific queries. Drawing on Media Richness Theory, we propose that chatbots represent a richer medium than static GenAI summaries. We also examine the definite valence of GenAI summaries as a moderator of their influence on purchase intention. Using eye-tracking technology, we measure cognitive engagement during a simulated e-commerce shopping task. Metrics such as fixation duration, fixation count, and time to first fixation across predefined Areas of Interest (AOIs) will be analyzed to understand how consumers process GenAI tools and their joint impact on decision making.

5 - Causal Modeling of User Retention with Large Language Models for Decision-Driven Platforms

YITING TANG, Univeristy of Sydney, Sydney, Australia

We develop a unified causal modeling framework for user retention in online platforms, combining generative language models with heterogeneous treatment effect estimation. As digital platforms increasingly rely on personalized engagement strategies, it becomes critical to understand not only *what* interventions work, but *why* and *for whom*. Our framework addresses this challenge by integrating psychological features extracted from user-generated texts—such as reviews and browsing histories—using large language models (LLMs), into a structural causal analysis of user churn.

We begin by encoding high-dimensional textual data into semantically meaningful embeddings, which serve as individualized covariates reflecting dissatisfaction, sentiment polarity, and engagement dynamics. These covariates are then incorporated into a potential outcomes framework, where treatment is defined as a time-sensitive promotional intervention (e.g., issuing coupons several days after first purchase), and outcome as the likelihood of user repurchase. To estimate causal effects, we employ causal forests to compute conditional average treatment effects (CATE), while using doubly robust estimation with cross-fitting to reduce selection bias from observational data.

We further analyze treatment effect heterogeneity across behavioral and demographic subpopulations, revealing actionable insights for targeted marketing strategies. Empirical validation on real-world e-commerce datasets demonstrates that incorporating LLM-derived psychological signals improves AUC in churn prediction and increases policy value compared to conventional models. These results underscore the potential of integrating generative AI with causal inference frameworks to drive interpretable, adaptive, and demographically personalized retention strategies in decision-driven platforms.

SC61

Building B Level 4 B401

QSR Student Introduction and Interaction Session

Panel Session

Quality, Statistics and Reliability

Co-Chair: Alan Vazquez, Tecnologico de Monterrey, Monterrey, 64700, Mexico

Co-Chair: Yu Jin, University at Buffalo, Binghamton, NY, 14260, United States

1 - Moderator Panelist

Yu Jin, University at Buffalo, Binghamton, NY, United States

This interactive session serves as a vital platform for connecting PhD students with the broader Quality, Statistics, and Reliability community. Following the successful format of previous years, the session begins with a series of concise "elevator pitches" where selected QSR student members (especially those nearing graduation) present their research interests and accomplishments. This showcase segment provides valuable exposure for emerging talent. The second part features an interactive mini-panel discussion and Q&A with invited senior QSR faculty and industry professionals. These experienced members will share insights and advice on navigating the academic and industry job markets, developing research programs, teaching effectively, and building a successful career in QSR-related fields. The session aims to foster networking, highlight student research, and provide targeted mentorship.

2 - Moderator Panelist

Alan Vazquez, Tecnologico de Monterrey, Monterrey, Mexico

This interactive session serves as a vital platform for connecting PhD students with the broader Quality, Statistics, and Reliability community. Following the successful format of previous years, the session begins with a series of concise "elevator pitches" where selected QSR student members (especially those nearing graduation) present their research interests and accomplishments. This showcase segment provides valuable exposure for emerging talent. The second part features an interactive mini-panel discussion and Q&A with invited senior QSR faculty and industry professionals. These experienced members will share insights and advice on navigating the academic and industry job markets, developing research programs, teaching effectively, and building a successful career in QSR-related fields. The session aims to foster networking, highlight student research, and provide targeted mentorship.

3 - Panelist

Raed Al Kontar, University of Michigan, Ann Arbor, MI, United States

4 - Panelist

Kamran Paynabar, ISyE Georgia Tech, Atlanta, GA, United States

5 - Panelist

Shan Ba, LinkedIn, San Francisco, CA, United States

6 - Panelist

Wenmeng Tian, Mississippi State University, Mississippi State, MS, United States

SC62

Building B Level 4 B402

2025 PGE Energy Analytics Challenge: Solutions and Lessons Learned

Panel Session

Quality, Statistics and Reliability

Co-Chair: Murat Yildirim, Wayne State University, Detroit, MI, United States

Co-Chair: Ahmed Aziz Ezzat, Rutgers University, Piscataway, NJ, United States

1 - Moderator Panelist

Ahmed Aziz Ezzat, Rutgers University, Piscataway, NJ, United States

The panel will discuss the findings and lessons learnt from the 2025 PG&E Energy Analytics Challenge, co-organized by IISE and INFORMS, and sponsored by the Pacific Gas & Electric Company (PG&E) -- one of the largest electric utilities in the U.S. The 2025 challenge theme was on electric load forecasting, where contestants were provided multi-year electric load data, together with exogeneous weather and operational information. The competition challenged contestants to design data-science-based forecasting models to predict day-ahead electric loads. This panel will first introduce the challenge, as well as an overview of the methods used by the top-performing contestants. A distilled summary of the key insights and lessons learned by the challenge organizers will then be interactively discussed with the contestants, PG&E representatives, and the audience, as well as their relevance to the field of energy analytics and forecasting.

2 - Moderator Panelist

Murat Yildirim, Wayne State University, Detroit, MI, United States

3 - Panelist

Zekai Fan, Carnegie Mellon University, Pittsburgh, PA, United States, Shixiang (Woody) Zhu

4 - Panelist

Erick Jones, University of Texas at Arlington, Arlington, TX, United States

5 - Panelist

Yicheng Wang, Georgia Institute of Technology, Atlanta, GA, United States

6 - Panelist

Yu Zhang, University of California, Santa Cruz, Santa Cruz, CA, United States

7 - Panelist

Millend Roy, Columbia University, New York, NY, United States

SC63

Building B Level 4 B403

Data-Driven Reliability and Maintenance

Contributed Session

Contributed

Chair: Mobolaji Shobanke, George Washington University, Washington, DC, United States

1 - An Interleaved-Deletion Markov Boundary Discovery Method for Key Process Feature Identification in Manufacturing Systems

Jiaqi Zhang, Tianjin University, Tianjin, China, People's Republic of, Liangxing Shi

Identifying key process features (KPF) is critical to ensure product quality in manufacturing systems. The task is complicated by the large number of quality features and the underlying mechanisms among features, which often hinder accurate KPF identification. To tackle this challenge, this study proposes an improved Markov Boundary discovery method, called Interleaved-Deletion Markov Boundary (IDMB), for causal feature selection in manufacturing quality control. The IDMB method introduces three key innovations. First, it allows overlap between candidate parent-child (PC) and spouse node sets, broadening the search space for causal features and improving completeness. Second, it employs an interleaved deletion mechanism that removes potential false-positive spouses one PC node at a time. It reduces the risk of discarding true causal features erroneously due to conditional independence test errors. To cope with the problem of product quality data imbalance, IDMB allows embedded sampling methods. The proposed method extensively validates IDMB on eight standard Bayesian network benchmarks and six real-world manufacturing quality datasets from the UCI repository. The results show that IDMB achieves higher recall in identifying true causal features, outperforming thirteen state-of-the-art Markov Boundary discovery methods and four non-causal feature selection methods. The proposed approach advances the methodology of causal feature selection by improving identification accuracy and robustness and offers practical value for manufacturing quality control.

2 - Reliability Optimization of Air Traffic Control System

Merve Uzuner, Baskent University, Ankara, Turkey, Orhan Dengiz, Berna Dengiz

Communication networks within air traffic control (ATC) systems play a crucial role in ensuring secure and reliable communication among aircraft, ground control units, and other various air traffic management units. These networks enable the transfer of critical information such as flight plans, position reports, meteorological data, and control instructions. A wide range of technologies, such as satellite communications, data linkages, and radio frequencies, are used to provide coverage, particularly in distant and transoceanic areas. Communication networks are made up of three main components: ground-based communication stations, satellite systems, and onboard communication devices. These components work together in an integrated manner using complicated communication protocols and operational procedures. Due to increased air traffic, enhancing the reliability of these networks has become a significant area of research for the sustainability of air traffic management systems.

Efforts to improve system reliability are typically addressed in the literature under the framework of the "reliability optimization problem." The Redundancy Allocation Problem (RAP) is frequently used to solve this issue. This paper formulates and addresses the design of a highly reliable communication network within an ATC system using RAP. To solve RAP, which is known to be an NP-hard problem, a well-established heuristic method—Genetic Algorithm (GA)—is proposed. The algorithm's performance is evaluated through computational analysis based on varying problem sizes.

3 - Swinnext-Haada: Hierarchical Adaptive Attention-Driven Dual-Stream Architecture for Medical Image Classification

Tien Truong, University of South Florida, Tampa, FL, United States, Phat Huynh, Ca Hoang, Khanh Do, Son Pham, Han Lam, Trung Le

Deep learning methods have facilitated significant progress in medical image classification; however, substantial challenges remain in unifying anatomically global context with locally detailed features and adaptively selecting salient patterns across heterogeneous disease presentations. We propose *SwinNeXt-HAADA*, a dual-stream architecture that integrates Swin Transformer and ConvNeXt backbones for end-to-end medical image analysis. Our framework features a systematic, cross-stream fusion of *Dynamic Attention Refinement (DAR)* and *Stage Importance Weighting (SIW)*: DAR selectively amplifies relevant channels at the fusion level, while SIW highlights cross-stage combinations that yield the greatest discriminative power. These mechanisms enable *SwinNeXt-HAADA* to dynamically emphasize both large-scale contextual cues (via transformer-based global attention) and fine-grained local details (via convolution-driven feature extraction), even in anatomically complex images. We evaluated our approach on three medical imaging domains—brain tumor MRI, chest X-ray, and gastrointestinal endoscopy—totaling over 22,000 images. The proposed model achieved up to 99.56% accuracy for brain tumor classification, 98.70% for chest disease detection, and 98.53% for endoscopic lesion recognition, surpassing state-of-the-art methods. Further analysis of the learned attention distributions revealed that channel-level gating uncovers subtle but critical features (e.g., tumor boundaries), while stage-level weighting adaptively prioritizes synergy between deeper stages of both streams for MRI and mid-to-late or early-late pairings for other pathologies. Feature activation visualization underscored the effectiveness of a joint global-local strategy. Altogether, these results suggest that *SwinNeXt-HAADA* can serve as a strong candidate for multiscale, heterogeneous medical image classification.

4 - Computing Distributionally Robust Preventive Maintenance Intervals

Heraldo Rozas, Department of Electrical Engineering, University of Chile, Santiago, Chile, Nagi Gebrael, Weijun Xie

Preventive maintenance (PM) intervals are traditionally determined using predict-then-optimize (PTO) frameworks, wherein historical failure time (FT) data is first used to characterize the underlying FT distribution, and then an optimization problem is solved to determine the optimal PM intervals. Nonetheless, such PTO frameworks may perform poorly out-of-sample if the fitted FT distribution differs significantly from the true distribution. This talk addresses this issue by introducing a contextual Distributionally Robust (DR) model for computing optimal PM intervals. The proposed model integrates empirical FT data directly into the PM optimization framework without assuming any specific distribution. Potential perturbations of the empirical FT distribution are captured using a Wasserstein ambiguity set. The DR model admits tractable mixed-integer linear programming reformulations in various practical settings, making it easy to implement with off-the-shelf optimization solvers. Its efficacy is demonstrated through computational studies involving both simulated and real-world failure time data.

5 - Neighborhood-Based Modeling and Prediction of Spatiotemporal Melt-Pool Dynamics

Siqi Zhang, Southern Illinois University Edwardsville, Edwardsville, IL, United States, Hui Yang

Laser powder bed fusion (LPBF) is a metal-based additive manufacturing process to produce 3D parts directly from digital designs. In LPBF, a laser beam scans a powder bed to selectively fuse the material into the shape of the desired part. At each fusing point, the laser heats and liquidizes the powder, during which a melt pool is formed. The dynamics of the melt pool are the result of laser-powder interactions over time and significantly influence the microstructure and mechanical characteristics of the final part. Accurate prediction of melt pools is essential to many applications, e.g., optimizing process parameters and enabling real-time defect prevention. However, this prediction problem is challenged by spatial-temporal effects and time-varying process parameters in actual manufacturing. In this work, we develop a neighbor-based encoder-decoder framework to capture spatiotemporal effects and predict melt-pool dynamics. Specifically, the encoder consists of 1) a process module that captures the spatial heterogeneity between a melt pool and its neighbors due to process variations and 2) a temporal module to model time-related dependences within the neighboring region. Given a set of neighboring melt-pool images and their associated process variables, the encoder generates a spatiotemporal latent representation, which is then decoded for prediction. The proposed framework is evaluated using melt-pool imaging data from LPBF benchmark experiments. Compared to traditional spatiotemporal models, it not only incorporates process variables into the modeling but also accommodates non-continuous images as temporal data. This framework is broadly extensible to other problems involving spatiotemporal dynamics with non-continuous data.

6 - Timescale Selection and Reliability Modeling for Condition-Based Maintenance of Medical Equipment

Mobolaji Shobanke, The George Washington University, Washington, DC, United States, Connor Roberts, Ekundayo Shittu

Driven by growing interest in data-driven approaches to equipment reliability, the transition from fixed-interval preventive maintenance to condition-based maintenance (CBM) has gained traction in healthcare technology management. While recent studies focus on stochastic modeling techniques, the fundamental choice of degradation timescale is understudied. Most reliability models default to recorded calendar age in service logs. Potentially obscuring relevant operational variability in systems where

usage intensity varies widely across units. This study presents a comparative analysis of reliability models using calendar time and stress-related metrics to predict the risk of failure in 121 repairable medical devices. Loglinear nonhomogeneous Poisson process (NHPP) with inverse gaussian frailty models were developed for each timescale to estimate expected failures and assess predictive performance based on individualized equipment usage duration. Although calendar time model produced lower mean squared error in some equipment, it exhibited minimal prediction variance, and nearly uniform frailty estimates across the equipment in the study. In addition, it showed a tendency to regress toward the system-wide average failure rate; thus, indicating poor sensitivity to operational differences. Conversely, utilization-based models captured broader frailty distributions, higher prediction variance, and stronger correlation with observed failure patterns. These findings underscore the importance of operational stress measures as more appropriate indicators of equipment degradation and highlight how timescale selection fundamentally affect model fidelity. This empirical validation provides practical evidence to theoretical CBM frameworks and offers guidance for practitioners implementing condition-based maintenance strategies in systems with differentiated usage.

SC64

Building B Level 4 B404

Human Decisions and Technological Interfaces: Behavioral and Operational Insights in Modern Systems

Invited Session

Behavioral Operations Management

Chair: Lai Wei, Boston College, Chestnut Hill, MA, United States

1 - The Impact of Smart Micromobility on Urban Crime: Evidence from Bike-Sharing Systems

Wen Ya Shen, Arizona State University, Tempe, AZ, United States, Zenan Zhou, Jiding Zhang, Pei-yu Chen

Urban safety remains a preeminent concern for residents, profoundly influencing their quality of life, shaping business activities, and guiding urban planning initiatives. Smart micromobility systems are emerging technology-based solutions that leverage digital platforms and infrastructure to provide short-term access to lightweight shared vehicles. However, there is ongoing debate over whether these systems enhance public safety or inadvertently contribute to crime. On the one hand, bike-sharing can improve transit accessibility and increase street activities, potentially enhancing mobility and fostering natural surveillance, which may deter crime (the crime-reducing effect). On the other hand, offenders may exploit the convenience of shared bikes and designated parking areas to facilitate criminal activities (the crime-inducing effect). This study examines the impact of smart micromobility systems on urban crime and the underlying mechanisms. Building on crime activity theories and the Land Use and Transportation (LUT) Planning framework, we conceptualize smart micromobility as an intermediary between automobility (including transit mobility) and pedestrian mobility. By increasing speed and mobility for potential targets, reducing exposure to offenders, and providing flexibility to circumvent high-risk areas, these systems may influence crime patterns. Leveraging a large-scale dataset on bike-sharing deployment in Chicago, we find that the introduction of bike-sharing has an overall crime-reducing effect. The effect is more pronounced in areas with higher automobility constraints and greater transit accessibility. Additionally, our analysis of crime patterns and neighborhood heterogeneity reveals that neighborhoods with higher reliance on smart micromobility and those experiencing opportunistic crimes benefit more significantly from the system's crime-reducing effect.

2 - Replacing What Could be Repaired: a Structural Analysis of Two-Stage Diagnostic Decisions in Managing Shared-Bike Returns

Jingxuan Geng, University of Michigan, Ann Arbor, MI, United States, Hailong Cui, Guangwen Kong

Bike-sharing platforms face significant challenges from high maintenance costs, driven by heavy usage and inefficiencies in diagnostic decision-making. Using task-level data from a leading bike-sharing platform, we develop a structural estimation model to analyze two-stage diagnostic decisions made by inspectors (stage 1) and workers (stage 2). These decisions are modeled as a strategic interaction governed by a Bayesian Nash Equilibrium (BNE). To address the computational complexity of Maximum Likelihood Estimation with BNE constraints, we employ machine learning to approximate BNE. We identify systematic overtreatment tendencies among inspectors and workers, resulting in a higher false positive rate than that under the firm's optimal decisions and thus inflating maintenance costs. Our counterfactual analyses show that higher part costs, reducing workers' piece-rate wages, adopting structured job matching, and prioritizing worker training can substantially reduce costs. Transitioning from a two-stage to a one-stage process lowers diagnostic accuracy and increases costs, although optimizing wages narrows this gap. This framework provides actionable insights for mitigating inefficiencies in multi-agent diagnostic decision systems and is generalizable to other credence goods industries, such as heavy equipment maintenance and healthcare, where diagnostic errors have significant financial, operational, and health implications.

3 - Stress Trauma and Decision Making in Humanitarian Logistics

Julie Niederhoff, Syracuse University, Syracuse, NY, United States, Llord Brooks

Experience and trauma may influence humanitarian aid workers' decision-making in inventory management and resource allocation. Through a behavioral experiment, we compare decisions across varying levels of trauma and experience to identify when decision-makers are most vulnerable to new information and potential trauma triggers. We also examine which factors (e.g., speed, quantity, cost) they prioritize during the decision-making process.

4 - Hagglng with a Bot: Human vs. LLM Negotiation in Supply Chain Contracts

Rihuan Huang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Yang Chen

Buying and selling by autonomous AI agents are no longer visions of the future. Large companies like Walmart employ AI agents in contract negotiations. However, AI-human negotiation in the supply chain context is not well understood. In this paper, we experimentally investigate the negotiation dynamics and outcomes of supply chain contract terms between human and large language model (LLM) agents. Specifically, we consider a two-tier supply chain comprising an upstream supplier and a downstream retailer. The retailer purchases units from the supplier at a per-unit wholesale price and fulfills random demand. We design an experiment in which participants act as retailers and negotiate the wholesale price and quantity with an LLM-powered supplier. Our experiment follows a 2×2 design, varying two key dimensions: 1) which party, retailer or supplier, bears the cost of unsold inventory, and 2) whether the retailer can engage in free-form conversations during negotiations. The experimental results show that, compared to human-human negotiations, suppliers represented by LLM agents achieved similar bargaining outcomes when they bear the risk of unsold inventory. In contrast, when retailers take on this risk, the LLM agents' performance depends on the negotiation format. When free-form conversations are enabled, LLM agents perform similarly to humans. However, when free-form conversations are disabled, LLM agents significantly outperform humans by increasing supplier profit through higher wholesale prices, while the channel profit remains unchanged.

5 - Large Language Models in the Workplace: a Social Perspective

Lai Wei, Boston College, Chestnut Hill, MA, United States

Powered by Large Language Models (LLMs), AI tools such as ChatGPT provide a compelling alternative to seeking guidance from human supervisors when workers face challenges in the workplace. However, adoption of these general-purpose tools in the workplace can be affected by workers' prior personal experiences with LLMs which are often influenced by workers' social attributes such as social class. This leads to unequal adoption of LLMs across different social groups. Using surveys and controlled experiments, we examine the differences in the adoption of LLMs as a substitute for interactions with human supervisors among workers from lower, middle, and upper social classes, measured through both objective and subjective dimensions. For objective social class, defined by parental education level and income, both lower- and upper-class participants are less likely to adopt LLMs compared to middle-class participants. For subjective social class, based on participants' perceptions of their parental social rank, upper-class participants also show a lower preference for LLMs than the middle class, which is driven by their lower social interaction costs when seeking guidance from supervisors. Interestingly, while lower-class participants generally exhibit similar preferences to the middle class, those with limited understanding of LLMs' capabilities, limitations, and application contexts, are less likely to adopt LLMs compared to the middle class. This paper provides important insights for workers, companies, and policymakers aiming to promote equitable and effective integration of LLMs in the workplace.

SC65

Building B Level 4 B405

AI in Healthcare

Invited Session

Artificial Intelligence

Chair: TJ Guo, Kelley School of Business, Bloomington, IN, 47405, United States

Co-Chair: Dezhi Wu, University of South Carolina, 550 Assembly Street, Columbia, SC, 29208, United States

1 - Fairness Evaluations on Continuous Glucose Monitor Large Sensor Models

Junjie Luo, Johns Hopkins University, Baltimore, MD, United States

Foundation models are becoming central to AI applications in healthcare, with large sensor models (LSMs) trained on continuous glucose monitoring (CGM) data demonstrating strong predictive capabilities for diabetes management. Despite their promise, little is known about whether these models perform fairly across diverse patient populations. In this study, we conduct a fairness evaluation of CGM-based LSMs by examining performance disparities across diabetes type, age, and gender groups. Using near-future glucose prediction as the benchmark task, we uncover systematic gaps in accuracy and calibration, particularly for younger individuals, females, and those with Type 1 diabetes. These disparities raise concerns about equitable model deployment and highlight the need for fairness-aware development and evaluation practices to ensure that LSMs benefit all patient groups reliably.

2 - Using Latest Large Language Models and Prompting Strategies for Social Media Analysis: a Feasibility Study for a use Case in Healthcare

Anthony Fu, USC and DFHS, Columbia, SC, United States

Youth vaping poses serious health risks, and social media offers key insights into user behavior. This study evaluates the capacity of the latest large language models (LLMs)—including GPT-o1, GPT-o3-mini, Gemini 2.0 Flash, Gemma 2, LLaMa 3.3, R1, and Grok-2—to detect user quit intentions and cessation stages from social media posts. Model outputs were benchmarked against human annotations using various prompting strategies. Preliminary results indicate OpenAI's GPT-0.1 is the best-performing model, followed by LLAMA 3.3; the worst was Gemma 2. Although the current preliminary results demonstrate the potential of LLMs, none of them can replace human annotators.

3 - Designing An AI-Driven Clinical Decision Support Systems for Hiv Care

Dezhi Wu, University of South Carolina, Columbia, SC, United States

This work-in-progress study aims to apply "human-in-the-loop" design strategies to develop and implement an AI-driven clinical decision-support system for HIV clinicians. Driven by a large population-level HIV patient electronic health record dataset from a southeastern state in the US, we aim to use clinical machine learning models to classify different levels of HIV patient risks and provide AI-driven recommendations to HIV clinicians for more informed treatment decisions. Using a mixed-methods approach, this study will provide timely design implications for incorporating AI outputs to support HIV clinical decision support processes. More results will be updated before the INFORMS 2025 conference.

SC66

Building B Level 4 B406

Pedagogical Design and Innovations for OR and Supply Chain Education

Invited Session

INFORMS Committee on Teaching and Learning

Chair: Yulia Vorotyntseva, University of St Thomas, Minneapolis, MN, United States

1 - Teaching Retail Assortment Strategy: the Assortment Planning Game Simulation

Yulia Vorotyntseva, University of St Thomas, Minneapolis, MN, United States

This work presents a web-based simulation in which learners engage in retail assortment planning — that is, selecting which products to offer in a given category. Market demand is modeled using the Multinomial Logit (MNL) consumer choice model -- a widely used framework in assortment optimization -- making the simulation a practical tool for introducing this model to learners. The simulation is implemented on the SoPHIELabs Classroom platform.

Through this interactive game, learners explore the trade-offs between expanding market share by offering a greater variety of products and reducing operational costs through variability pooling. It enables them to experientially grasp and internalize key concepts such as the cost implications of increased product variety (including higher inventory and overhead costs), and the phenomenon of product cannibalization (where additional variety may not significantly increase revenue).

By engaging with this simulation, learners connect theoretical models with practical decision-making skills essential for managing retail operations.

2 - Continuous Improvement for Impact: Case Development, Industry Collaboration, and Student Engagement

Xiao Zhang, Saint Louis University, Saint Louis, MO, United States, Varun Gupta

This session aims to share practical insights and strategies for developing meaningful business case studies, fostering industry research partnerships, and engaging students in impactful ways. Using multiple case studies as examples, we will guide participants through case development processes, publishing options, and collaboration opportunities.

3 - Contributing to a More Sustainable World One Student at a Time for over a Decade

Tarkan Tan, University of Zurich, Zurich, Switzerland

In 2013, we developed a course on sustainable supply chains that introduces key concepts and decision-making challenges in environmentally and socially responsible operations and supply chains, well before the topic became popular in business education. A core element of the course is a real-life project carried out by student teams in collaboration with companies or organizations. This challenge-based learning approach allows students to apply what they learn to practical sustainability issues. The course also adopts a flip-the-classroom format: students present selected topics and project progress to their peers, receiving structured feedback and contributing to one another's learning. Peer assessment is used as part of the grading process to encourage active engagement and critical reflection. To support the course, we co-edited a textbook -now in its second edition- that combines research-based insights with practical relevance. While originally designed for master's students, the course has been adapted to bachelor's, PhD, and executive education settings by adjusting content depth, project expectations, and delivery style.

This talk draws on over a decade of experience in designing and delivering the course. We will share reflections on its evolution, how it can be tailored to different audiences, and how it has helped students bring sustainability thinking into their careers, some becoming ambassadors for responsible business practices, others working directly on sustainable operations after graduation.

4 - Compete OR Collaborate? Experimental Insights on Team-Based Learning in Predictive Analytics Education

Nicolas Nunez, Centrum PUCP, Lima, Peru

Our study explores how team configuration—collaborative vs. competitive—affects student learning outcomes in a predictive analytics course. Students were assigned to work in teams under two distinct pedagogical conditions using a controlled experimental design within a graduate-level classroom. One emphasized cooperation, the other gamified competition, while both followed the same course structure and tasks. Preliminary results suggest that while competitive settings may boost engagement and speed, collaborative teams tend to outperform in depth of understanding and model interpretation. This work contributes to designing more effective data science education strategies by aligning pedagogical methods with cognitive and behavioral outcomes.

SC67

Building B Level 4 B407

AI and Machine Learning for Modeling, Prediction, and Decision Making in Healthcare

Invited Session

Data Mining

Chair: Jianxin Xie, University of Virginia, Charlottesville, VA, United States

Co-Chair: Xin Zan, University of Iowa, Coralville, IA, United States

1 - Quantile-Based Monitoring of Profiles with Generally Distributed Responses

Xin Li, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Jian Li, Kai Wang

Conventional profile monitoring employs linear regression for modeling, which investigates how the mean of the response depends on explanatory variables. This, however, renders such methods highly constrained by the assumption of normally distributed responses. In contrast, quantiles that are derived from the concepts of sorting and ranking make no assumptions about the response distribution. We employ quantile regression to characterize the functional relationship between the response quantile and explanatory variables. And by jointly monitoring multiple different quantile regression coefficients, we are in a position to monitor the relationship between the entire distribution of the response and explanatory variables. Furthermore, to address the computational complexity associated with parameter estimation in quantile regression, a Hotelling's T^2 type monitoring statistic is proposed by leveraging the Bahadur representation of conditional quantiles and their asymptotic distribution. Extensive simulations have demonstrated the superior performance of the proposed method across various types of response distributions.

2 - Transformers with Koopman-Enhanced Graph Convolutional Network for Spatiotemporal Dynamics Forecasting

Zekai Wang, Fairfield University, Fairfield, CT, United States, Bing Yao

Spatiotemporal forecasting is essential for understanding and controlling dynamic systems in fields ranging from transportation to biomedicine. However, modeling these processes remains challenging due to irregular spatial geometries, nonlinear dynamics, and long-range temporal dependencies. In this study, we propose a novel two-stage predictive framework, Transformer with Koopman-enhanced Graph Convolutional Network (T-KGCN), to address these challenges. Our approach integrates graph convolutional networks (GCNs) to model complex spatial structures, Koopman operator theory to embed the system's state into a latent space with approximately linear dynamics, and a transformer to capture long-range temporal dependencies. Our contributions include: (1) a unified architecture combining GCN, Koopman embedding, and transformers; (2) a principled approach to model nonlinear dynamics with latent linear evolution; and (3) enhanced performance in long-range forecasting. While demonstrated in a cardiac setting, our framework is broadly applicable to diverse spatiotemporal prediction tasks.

3 - A Self-Supervised Self-Corrected Masked Autoencoder with Brain Tumor Applications

Hairong Wang, Georgia Institute of Technology, Atlanta, GA, United States

Intra-tumoral heterogeneity poses a significant challenge for glioblastoma (GBM) diagnosis and treatment. Histopathological and molecular analyses are limited by surgical sampling, highlighting the need for non-invasive methods to map these alterations across the entire tumor. Limited biopsy samples hurdle the power of artificial intelligence (AI) models, and patient heterogeneity further complicates the task with small sample sizes. We propose a self-supervised framework using abundant MRI data beyond biopsy regions to enhance AI model robustness and generalizability. Our framework consists of two steps. First, we train generative models in a self-supervised manner to translate each patient's image into a synthetic "healthy" version, under the hypothetical condition that the patient does not have the tumor, through vision transformer based masked autoencoder. This step captures the inherent differences in patients' brains. Second, we classify regional biomarkers such as hallmark genes, tissue states, and cell types based on both real and generated MRI. Our study uses 283 image-localized biopsies with corresponding T1Gd and T2 MRI data from 67 GBM patients. Our model achieved satisfying classification accuracy, surpassing state-of-the-art methods. Additionally, we generated biomarker prediction maps across entire tumors, providing a quantitative assessment of intra-tumoral heterogeneity in histopathological and molecular alterations. These results highlight the emerging role of AI in Precision Medicine and its potential for broader applications in GBM diagnosis and treatment.

4 - Heterogeneous Domain Adaptation for Survival Analysis

Dan Ni Lin, Texas Tech University, Lubbock, TX, United States

Artificial Intelligence (AI) technologies have advanced clinical risk prediction by effectively handling complex data and uncovering insights from large medical datasets. However, small medical centers with limited and unlabeled patient data face challenges in adopting these tools, which rely on large, labeled datasets for development. Domain adaptation technique adapts models developed from larger centers or national registries to smaller healthcare settings by leveraging the knowledge and features learned from a large dataset to enhance generalization. However, differences in feature distribution and dimensionality between datasets pose challenges for existing domain adaptation methods. This study introduces a heterogeneous domain adaptation (HDA) method which is a semi-supervised approach to adapt neural network Cox Proportional Hazard models derived from large source datasets to smaller target datasets. The method integrates dictionary learning and adversarial training to address structural and statistical differences between source and target domains.

SC68

Building B Level 4 B408

Data Mining for Education and Learning Analytics

Invited Session

Data Mining

Chair: Ishita Malhotra, 75204

1 - Course Recommendation Using Domain-Based Cluster Knowledge and Matrix Factorization with Generative Ai-Driven Synthetic Data Augmentation

Satyam Mittal, California State University, Fresno, Dallas, CA, United States, Ishita Malhotra

We propose an enhanced course recommendation system combining domain-based clustering and matrix factorization, augmented with Generative AI for synthetic data curation. Unlike prior work that relies solely on historical enrollment data, our approach uses Generative AI to generate school-specific synthetic student-course interactions, tailored to academic domains and performance trends. This enriched data enables robust training even for institutions with sparse records, improving personalization, fairness, and privacy in educational recommendations.

2 - Tracing Epistemic Influence Through Semantic Embedding Spaces

Onur Seref, Virginia Tech, Blacksburg, VA, United States

This study explores how fields of knowledge evolve by modeling changes in their underlying semantic structures over time. Using a large corpus of research abstracts from top Information Systems journals, we construct a temporally layered embedding space to investigate how influential contributions relate to local shifts in meaning, including conceptual drift, turbulence, and re-centering. We combine entropy-based field metrics, clustering coherence, and citation-informed dynamics to detect signals of epistemic movement across time and context. This work offers a preliminary step toward AI-assisted field analysis, where semantic embeddings can help characterize, visualize, and ultimately forecast how knowledge systems evolve, reposition themselves, or reorganize across conceptual or institutional boundaries.

3 - Feddes: Graph-Based Dynamic Ensemble Selection for Personalized Federated Learning

Brianna Mueller, University of Iowa, Iowa City, IA, United States

Dynamic ensemble selection (DES) methods aim to enhance predictive performance by identifying a subset of classifiers most suitable for each individual data sample. These methods often assume that samples in close proximity exhibit similar classification behavior and therefore estimate classifier competence within a local

neighborhood of the query. However, sensitivity to local structure can limit robustness and applicability, as traditional DES techniques may overlook complex interdependencies between classifiers and data samples. To address this, we propose a novel graph-based DES framework that leverages graph neural networks (GNNs) to learn sample-specific ensemble weights. Our approach is particularly motivated by the challenges of personalized federated learning (pFL), where clients possess diverse local data distributions and benefit from tailored model selection. In the proposed method, each client constructs a heterogeneous graph in which nodes represent local data samples and classifiers shared by other participants in the network, while edges encode relationships such as classifier confidence, sample similarity in the decision space, and classifier diversity. During inference, test samples are inductively inserted into the graph, allowing the model to leverage learned relationships to output classifier weights for a personalized ensemble prediction. We evaluate our method on the eICU Collaborative Research Database, treating each hospital as a federated client and each locally trained model as a candidate classifier. Our results demonstrate that modeling sample-classifier relationships through a graph-based framework can enhance generalization for personalized federated learning.

4 - Optimization-Based Dimensionality Reduction

Deniz Yoltay, Ozyegin University, Istanbul, Turkey, Erhun Kundakcioglu

Dimensionality reduction is a fundamental task in data visualization, aiming to represent high-dimensional data in a low-dimensional space while preserving essential structural relationships. This study introduces a novel optimization-based framework that formulates the problem using a set of distance-preserving constraints. The proposed approach aims to capture key relational patterns among data instances and maintain geometric fidelity in the resulting low-dimensional embedding.

SC69

Building B Level 4 B409

Healthcare Innovation: Business Models and Technology

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Ersin Korpeoglu, UCL School of Management, London, United Kingdom

Co-Chair: Yupu Sun, University College London, London, United Kingdom

1 - Managing Prediction Errors in Healthcare Capitation: a Two-Stage Risk Adjustment Approach

Zhaowei She, Lee Kong Chian School of Business, Singapore Management University, Singapore, Singapore, Kanix Wang, Yajing Chen

This study tackles a fundamental challenge in capitation-based healthcare payment: prediction errors in risk adjustment that result in systematic underpayment for some patients and overpayment for others. These errors compromise both insurer budget neutrality and the actuarial soundness of the system. Motivated by a recent reform proposal from the Centers for Medicare & Medicaid Services (CMS), we develop a novel two-stage estimation algorithm for the HHS-HCC risk adjustment model. Our method directly targets and corrects persistent prediction errors—especially those that incentivize patient selection at the extensive margin. Using administrative claims data from the ACA Marketplace, we empirically show that our approach substantially improves predictive accuracy and achieves near-perfect actuarial equivalence across patient risk groups.

2 - Doctor + AI: Rethinking Medical Communication for the Digital Age

Mehmet Ayyaci, The University of Texas at Dallas, Richardson, TX, United States, Dicle Yagmur Ozdemir, Alejandro Zentner

As AI tools increasingly mediate healthcare communication, an open question is how best to integrate them into physician workflows. While AI-generated responses are often perceived as more helpful, clear, and respectful than those of physicians, concerns remain about misinformation and loss of human oversight. This study evaluates whether AI-physicians collaboration can improve communication without sacrificing clinical credibility. To do so, we conducted an experiment where both laypeople and medical professionals evaluated medical responses across four conditions: (1) physician-only, (2) AI-only, (3) AI-refined physician responses, and (4) AI-physician combined responses. In refinement, AI improves tone without adding medical content; in combination, AI merges its response with a physician's, unseen. Our results show that both lay and expert evaluators prefer AI-only and AI-physician collaboration responses over physician-only ones. Laypeople show a stronger preference for AI, but AI-physician collaborations, particularly refinement, narrow this preference gap, aligning lay evaluations more closely with expert judgments. These findings suggest that hybrid communication models can enhance medical communication while maintaining expert oversight. Rather than replacing physicians, AI can be used to augment their communication, offering a pathway for more effective, trustworthy, and scalable doctor-patient interactions.

3 - The Role of Hospital Rcm Systems in Mitigating Personal Bankruptcy Filings

Shuhua Wu, Temple University, Philadelphia, PA, United States, Chenzhang Bao, Sezgin Ayabakan, Min-Seok Pang

Medical debt has become a challenging concern in the US. Its overall amount has climbed to \$220 billion nationwide, and nearly 41% of the adults currently owe medical debts, especially those with high medical needs and who belong to disadvantaged groups. Consequently, medical debts have become one of the leading causes for individual bankruptcy filings. This study investigates how revenue cycle management (RCM) systems adopted by hospitals affect personal bankruptcy filings. By improving information completeness and accuracy, smoothing information flow, and streamlining the billing and collection process, RCM systems can decrease the uncertainty of revenue collection and improve the providers' financial stability. This allows hospitals to find alternative financial protection for near-poor patients, decreasing their propensity to file for bankruptcy. On the other hand, the improved efficiency in the billing and collection processes could also exacerbate bankruptcy filings as near-poor patients without insurances usually face larger medical bills than insured patients. We draw on organizational information processing theory to understand why hospitals choose to adopt RCM systems and how it affects individual bankruptcy filings. We construct a panel dataset from multiple sources and employed two-way fixed-effect models with instrumental variables to address the endogeneity concerns. Our analysis shows that the average RCM system adoption rate by hospitals in a county is negatively associated with bankruptcy filings, especially when a state expands the Medicaid coverage. This study contributes to the literature that examines the societal impacts of information systems in healthcare settings and provides ample implications for practitioners and policymakers.

4 - Does Virtual Take Longer? An Empirical Investigation of Service Time in Telemental Health

Yupu Sun, University College London, London, United Kingdom, Ersin Korpeoglu

Telemedicine visits, in the form of synchronous virtual interactions between patients and providers, have witnessed a significant surge during the COVID-19 pandemic. In this study, we employ an instrumental variable model to examine the causal relationship between the form of psychology treatments (virtual or in person) and the service time of these treatments. Using visit-level patient data for psychology visits in the United States (US), we find that telemedicine use significantly increases cumulative service time and visit frequency per month. Depending on which of these effects dominates, the duration of single visits can decrease or increase when using telemedicine. Specifically, we find that telemedicine use decreases the duration of single visits for private insurance patients while increasing it for Medicare insurance patients. Furthermore, we find that telemedicine use leads to a larger increase in per-visit service time for retired patients than for non-retired ones. These results show that serving Medicare patients, especially retired ones, with telemedicine can substantially increase the burden on healthcare providers. Our study contributes to a better understanding of how telemedicine affects the efficiency of care delivery in mental health.

5 - The Rise of Subscription-Based Healthcare: Strategic Impacts on Patients and Providers

Ozden Cakici, American University, Washington, DC, United States, Hessam Bavafa

Subscription-based healthcare systems (SBHS) offer patients unlimited access to primary care for a fixed fee, e.g., Amazon One Medical, Blueberry Pediatrics. This revenue model challenges traditional fee-for-service (FFS) models that bill per visit. While SBHS are designed to reduce administrative burden and improve access, particularly for low-acuity patients, their expansion raises concerns about equity and market efficiency.

We develop an analytical model to examine competition between SBHS and FFS providers, capturing patient decisions based on illness severity, costs, and wait times. SBHS attract noncritical cases, relieving FFS congestion but introducing their own capacity constraints. Providers must set subscription fees and define the maximum severity they can manage to maintain profitability and service quality.

Our analysis shows that SBHS alter the structure of congestion rather than eliminating it. Equilibrium outcomes depend on the acuity distribution of the population and market conditions. Financial sustainability may require excluding certain patient groups, and even when both high- and low-acuity patients are included, welfare gains are unevenly distributed.

These findings challenge the assumption that SBHS universally improve access and equity. We offer guidance for healthcare managers and policymakers on integrating SBHS to enhance system performance while mitigating disparities.

Sunday, October 26, 2:45 PM - 4:00 PM

QSR Poster

Building A Level 1 Exhibit Hall A3

QSR Best Student Poster Competition

Award Session

Quality, Statistics and Reliability

Chair: Chao Wang, University of Iowa, Iowa City, IA, United States

Co-Chair: Mahathir Bappy, Mississippi State University, Mississippi State, MS, United States

1 - Poster - Automated Analysis of Experiments Using Hierarchical Garrote

Wei-Yang Yu, Georgia Institute of Technology, Atlanta, GA, United States

In this work, we propose an automatic method for the analysis of experiments that incorporates hierarchical relationships between the experimental variables. We use a modified version of the nonnegative garrote method for variable selection which can incorporate hierarchical relationships. The nonnegative garrote method requires a good initial estimate of the regression parameters for it to work well. To obtain the initial estimate, we use generalized ridge regression with the ridge parameters estimated from a Gaussian process prior placed on the underlying input-output relationship. The proposed method, called HiGarrote, is fast, easy to use, and requires no manual tuning. Analysis of several real experiments are presented to demonstrate its benefits over the existing methods.

2 - Poster - Generating Synthetic Aia Images from Historical Ground-Based Observations: a Diffusion Model Approach for Extended Coronal Hole Analysis

Marco Marena, New Jersey Institute of Technology, Newark, NJ, United States

Accurate and continuous observations of the Sun's extreme ultraviolet (EUV) emissions are critical for understanding solar dynamics and monitoring solar activity. However, EUV observations availability is limited to the past ~30 years, restricting our ability to study long-term solar corona. This study aims to reconstruct a historical dataset of synthetic AIA 193 Å and 304 Å — solar corona and chromosphere — using historical ground-based chromospheric observations in H α and CaII K, available for over a century. To generate AIA observations before 2010, we propose to use a novel deep learning methodology, the Diffusion Model Translation (DMT) framework, built on a fine-tuned Denoising Diffusion Probabilistic Models (DDPM), and a Two-Stream Image-to-Image Translation (TSIT) network as a translator between the noising and denoising processes, to bypass the diffusion steps and accelerate training. We used H α and CaII K lines from Kanzelhöhe Solar Observatory (KSO) spanning from 2011 to 2024, and AIA 193 Å and 304 Å from the Solar Dynamic Observatory (SDO). For the independent test set, the model shows a high full-disk correlation (CC = 0.942) and achieves a Dice coefficient of 0.893 for coronal holes. Particularly, the high Dice coefficient highlight the model's effectiveness in identifying and delineating coronal holes, sources of high-speed solar wind streams. By providing reliable synthetic EUV imagery for historical periods lacking direct observations, our methodology facilitates deeper analyses of coronal hole evolution, as well as long-term solar activity trends.

3 - Integrating Technological and Rights Attributes for Patent Portfolio Identification

Tongrui Zhang, Sun Yat-sen University, Guangzhou, China, People's Republic of

no abstract

4 - Poster - Hybrid Diffusion-Transformer Approach for Multivariate TIME Series Imputation

Ramiro Valdes-Jara, University of Miami, Miami, FL, United States

Many real-world spatiotemporal data contain gaps when sensors fail or communication issues occur. Spatiotemporal imputation focuses on reconstructing missing values by leveraging patterns that repeat across both space and time. In recent years, deep learning approaches have emerged as powerful tools for spatiotemporal imputation, consistently outperforming traditional methods. Although recent work has improved performance, two persistent gaps remain: (i) diffusion-based methods model uncertainty well but are confined to local temporal contexts and fixed spatial graphs, and (ii) transformer-style architectures capture long-range dependencies yet handle missing-value uncertainty only heuristically. We propose a hybrid method that addresses the limitations of each individual approach, namely, the Hybrid Diffusion Transformer (HDT). The diffusion component treats each time series as a noisy signal that can be gradually reconstructed, giving the model a way to handle uncertainty and to learn to remove the noise step by step. The transformer component adds an attention mechanism, enabling the model to simultaneously learn spatial correlations and non-stationary temporal patterns. During inference, the reverse diffusion trajectory, guided by the transformer component, iteratively reconstructs missing values while producing confidence intervals. By integrating these parts in a single, end-to-end framework, our approach offers a robust and interpretable solution for multivariate time series imputation. We evaluate HDT on varied traffic and physiological datasets chosen to test long-range and spatiotemporal dependencies. Model validation covers a range of missing-value patterns and compares mean absolute error (MAE) and root mean square error (RMSE) against state-of-the-art baseline methods.

5 - An Active Learning Strategy for Online Batch Fault Detection and Diagnosis

Yue Jiang, Purdue University, West Lafayette, IN, United States

No Abstract

6 - Learning to Detect the Unforeseen from Small Datasets: Manifold-Informed Meta-Learning for Out-of-Distribution

Zihan Li, University of Florida, Gainesville, FL, United States

no abstract

7 - Deep Belief Enabled Coupled High-Order State Transition Mode

Vipul Bansal, University of Wisconsin Madison, Madison, WI, United States

We propose a Deep Belief Updater (DBU) Model to address the challenge of modeling complex temporal dependencies in Coupled Higher-Order Hidden Semi-Markov Chains. These models extend Higher-Order Hidden Semi-Markov Chains by incorporating temporal dependencies among hidden states across multiple interdependent chains. Replacing underlying temporal dynamics with neural networks is computationally intensive, as it requires updating the state history at each time instance for each chain, based on the history of all the chains. To overcome this, we find that forward-state probabilities, the probabilities of being in a given state at a time instance based on the history of observations, can be expressed as a function of the Stage I belief states derived from individual observations. Stage I belief state is the likelihood of the system being in a certain state based solely on the current observation, without considering the influence of previous or future observations. The DBU model directly estimates these forward-state probabilities using Stage I belief. These probabilities are used to construct a likelihood function to train the model. By enabling the direct estimation of forward-state probabilities, the DBU model effectively captures the intricate transition dynamics of systems with interdependent chains.

8 - Airu-Wrf-Ens: A Multivariate, Time-Dependent Ensemble Method for Space-Time Wind Energy Forecasting

Jiaxiang Ji, Rutgers University, New Brunswick, NJ, United States

The reliable and cost-effective operation of offshore wind farms depends on accurate wind speed and power predictions across a range of forecast horizons. Yet, forecasting models typically perform well at a narrow range of look-ahead horizons: methods tailored for short-term forecasts (e.g., one or two hours ahead) rapidly degrade in accuracy as the forecast horizon extends (e.g., day-ahead), and vice versa. Using separate models for different forecast horizons is a possible solution, but would result in temporally discontinuous forecasts, limiting their value to wind farm and power system operations under uncertainty. In response, we propose an ensemble method which combines the outputs of multiple forecast models, each designed for a specific range of look-ahead horizons, thereby ensuring a temporally continuous forecast across the full forecast horizon. The proposed ensemble method preserves the temporal- and cross-correlations of individual ensemble members, introduces time-varying weighing to adapt to each model's degrading performance over time, and fully characterizes the predictive distribution of the ensemble output considering temporal- and cross-covariance information, enabling calibrated uncertainty quantification and decision support under uncertainty. Tested on offshore wind data from the Northeastern U.S.-a geographical region with significant offshore wind energy potential-the proposed approach achieves promising improvements, in terms of both point and probabilistic forecast evaluation metrics, relative to a representative set of forecasting methods.

9 - Distribution-in-Distribution-Out Regression on Pseudo Riemmanian Manifold of Empirical Measures

Yujing Huang, University at Buffalo, Buffalo, NY, United States

no abstract

10 - Hierarchical Influence via Spectral Shapley Active Learning (Hiss-Al) Strategy for in-Situ Nanosurface Morphology Characterizations Through Weak Sensor Signals

Xinchen Wang, Binghamton University, Vestal, NY, United States

no abstract

11 - Geopipn: Geometry-Aware Physics Informed Pointnet for Fast Thermal Distribution Prediction in Additive Manufacturing of Unseen Part Geometries

Amirul Islam Saimon, Virginia Tech, Blacksburg, VA, United States

no abstract

12 - Deep Reinforcement Learning and Combinatorial Optimization for Dynamic Workload and Maintenance Planning

Ying Fu, University of Wisconsin-Madison, Madison, WI, United States

no abstract

13 - Learning to Generalize: Dynamic Latent Variable Transfer Learning for Modeling Energy Demand in Industrial Systems

Afshin Asadi, University of Miami, Miami, FL, United States

no

abstract

14 - Udecode-VAE: Unsupervised Device-Level Energy Decomposition with Sequential Variational Autoencoders

Salih Salihoglu, University of Miami, Coral Gables, FL, United States

no abstract

15 - Poster - Tensor-Efficient High-Dimensional Q-Learning

Junyi Wu, University of Washington, Seattle, WA, United States

High-dimensional reinforcement learning faces challenges with complex calculations and low sample efficiency in large state-action spaces. Q-learning algorithms struggle particularly with the curse of dimensionality, where the number of state-action pairs grows exponentially with problem size. Current solutions include function approximation methods and neural network-based approaches like Deep Q-Networks, but these often suffer from instability and require extensive hyperparameter tuning. We propose Tensor-Efficient Q-Learning (TEQL), which applies low-rank tensor decomposition via block coordinate descent on discretized state-action spaces, incorporating novel exploration and regularization mechanisms. The key innovation is an exploration strategy that combines approximation error with visit count-based upper confidence bound to prioritize actions with high uncertainty, avoiding wasteful random exploration. Additionally, we incorporate a frequency-based penalty term in the objective function to encourage exploration of less-visited state-action pairs and reduce overfitting to frequently visited regions. Empirical results on large discrete control tasks demonstrate that TEQL outperforms conventional matrix-based methods and deep RL approaches in both sample efficiency and total rewards, making it suitable for resource-constrained applications, such as space and healthcare where sampling costs are high.

16 - Dynamark: a Reinforcement Learning Framework for Dynamic Watermarking in Industrial Machine Tool Controllers

Navid Aftabi, University of Washington, Seattle, WA, United States

no abstract

17 - Efficient Optimization of Plasma Radiation Detector Configurations Using Imperfect Inference Models

Difan Song, Georgia Institute of Technology, Atlanta, GA, United States

no abstract

18 - Poster - Deepoformer: Deep Operator Learning with Domain-Informed Features for Fatigue Life Prediction

Chenyang Li, New Jersey Institute of Technology, Fanwood, NJ, United States

Fatigue life characterizes the duration a material can function before failure under specific environmental conditions, and is traditionally assessed using stress-life (S-N) curves. While machine learning and deep learning offer promising results for fatigue life prediction, they face the overfitting challenge because of the small size of fatigue experimental data in specific materials. To address this challenge, we propose, DeepOFormer, by formulating S-N curve prediction as an operator learning problem. DeepOFormer improves the deep operator learning framework with a Transformer-based encoder and a mean L2 relative error loss function. We also consider Stüssi, Weibull, and Pascual and Meeker (PM) features as domain-informed features. These features are motivated by empirical fatigue models. To evaluate the performance of our DeepOFormer, we compare it with different deep learning models and XGBoost on a dataset with 54 S-N curves of aluminum alloys. With seven different aluminum alloys selected for testing, our DeepOFormer achieves an R2 of 0.9515, a mean absolute error of 0.2080, and a mean relative error of 0.5077, significantly outperforming state-of-the-art deep/machine learning methods including DeepONet, TabTransformer, and XGBoost, etc. The results highlight that our DeepOFormer integrating with domain-informed features substantially improves prediction accuracy and generalization capabilities for fatigue life prediction in aluminum alloys.

19 - Poster - Conformalized Decision Risk Assessment

Wenbin Zhou, Carnegie Mellon University, Pittsburgh, PA, United States

High-stakes decisions in domains such as healthcare, energy, and public policy are often made by human experts using domain knowledge and heuristics, yet are increasingly supported by predictive and optimization-based tools. A dominant approach in operations research is the predict-then-optimize paradigm, where a predictive model estimates uncertain inputs, and an optimization model recommends a decision. However, this approach often lacks interpretability and can fail under distributional uncertainty - particularly when the outcome distribution is multi-modal or complex - leading to brittle or misleading decisions. In this paper, we introduce CREDO, a novel framework that quantifies, for any candidate decision, a distribution-free upper bound on the probability that it is suboptimal. By combining inverse optimization geometry with conformal prediction and generative modeling, CREDO produces finite-sample risk certificates that are both statistically rigorous and practically interpretable. This framework enables human decision-makers to audit and validate their own decisions under uncertainty, bridging the gap between algorithmic tools and real-world judgment.

20 - Adaptive Sampling and Data-Efficient Monitoring of Spatiotemporal Data

Jiahui Zhang, University of Wisconsin-Madison, Madison, WI, United States

no abstract

21 - Characterizing Dynamic Spatiotemporal Patterns of Eeg for Emotion Recognition Using Weighted Heterogeneous Recurrence Network Analysis

Yujie Wang, University of Miami, Miami, FL, United States

no abstract

22 - Poster - Monitoring the Covariance of Multichannel Profiles

Davide Forcina, University of Naples Federico II, Naples, Italy

Modern industrial systems continuously produce multichannel profile data for real-time process monitoring and fault diagnosis. Although most existing techniques focus on detecting shifts in the process mean, changes in the covariance structure are equally critical, as the process behavior depends on the interactions among multiple variables.

In this work, we use functional graphical models to capture conditional dependencies in multichannel profile data, where high dimensionality and sparsity exacerbate model estimation and interpretability.

The proposed approach employs penalized likelihood ratio tests with adaptive penalty levels to detect various types of covariance changes.

Additionally, a diagnostic procedure based on change-point detection identifies which relationships have shifted. Extensive simulations and a real-world application involving multi-channel temperature profile monitoring confirm the proposed method's effectiveness over competing methods already appearing in the literature.

23 - Self-Expressive Federated Learning Framework

Mohammad Amini, University of Florida, Gainesville, FL, United States

no abstract

24 - Accelerated Brain Aging in Alzheimer's Disease with Lewy Body Co-Pathology: Insights from Brain MRI and Alpha-Synuclein SAAs

Babak Ahmadi, University of Florida, Gainesville, FL, United States

Alzheimer's disease (AD) and Lewy body (LB) pathology often occur together, but only recently has LB pathology become detectable in living patients via cerebrospinal fluid (CSF) α -synuclein seed amplification assays (SAA), enabling investigation of its joint impact with AD on neurodegeneration. We trained a deep learning brain age model on multi-cohort MRI scans from 4,355 cognitively unimpaired adults and applied it to 803 cognitively impaired participants. Individuals were assigned to four groups based on AD status ($p\text{-tau}_{181}/A\beta_{42}$) and LB status (SAA). The co-pathology group (AD+LB+) showed the greatest brain age acceleration. Model saliency maps indicated more widespread saliencies, which was corroborated by faster longitudinal atrophy across multiple regions and parallel cognitive decline. Together, these results suggest that LB pathology amplifies AD-related neurodegeneration and support the use of combined biomarker testing and targeted interventions for patients with co-existing AD and LB pathology.

25 - Poster - Adaptive Robust Optimization with Data-Driven Uncertainty for Enhancing Distribution System Resilience

Shuyi Chen, Carnegie Mellon University, Pittsburgh, PA, United States

Extreme weather events are placing growing strain on electric power systems, exposing the limitations of purely reactive responses and prompting the need for proactive resilience planning. However, existing approaches often rely on simplified uncertainty models and decouple proactive and reactive decisions, overlooking their critical interdependence. This paper proposes a novel tri-level optimization framework that integrates proactive infrastructure investment, adversarial modeling of spatio-temporal disruptions, and adaptive reactive response. We construct high-probability, distribution-free uncertainty sets using conformal prediction to capture complex and data-scarce outage patterns. To solve the resulting nested decision problem, we derive a bi-level reformulation via strong duality and develop a scalable Benders decomposition algorithm. Experiments on both real and synthetic data demonstrate that our approach consistently outperforms conventional robust and two-stage methods, achieving lower worst-case losses and more efficient resource allocation, especially under tight operational constraints and large-scale uncertainty.

26 - A Residual Guided Strategy with Generative Adversarial Networks in Training Physics-Informed Transformer Networks

Ziyang Zhang, New Jersey Institute of Technology, Newark, NJ, United States

no abstract

27 - Physics- and Image-Informed Reliability Prediction for Pipelines Under Competing Failure Modes

Nikesh Kumar, University of Arkansas, Fayetteville, AR, United States

no abstract

28 - Poster - Rampseg: a Robust Calibration-Free Image Segmentation Method for Reliable Melt Pool Analysis in Laser Powder Bed Fusion

Nazmul Hasan, University of Arizona, Tucson, AZ, United States

Coaxial monitoring of the melt pool, the molten region formed during laser exposure in laser powder bed fusion (LPBF) additive manufacturing, is critical for ensuring part quality, as its morphology reflects process stability. However, melt pool images are inherently stochastic due to variations in processing conditions, imaging setups, and transient phenomena such as spatter. The impact of this variability on segmentation accuracy has not been thoroughly explored in the literature, resulting in a critical gap in the development of robust methods. As a result, existing segmentation techniques often lack the generalizability and reliability needed to perform consistently across diverse conditions. In response, we propose RAMPSeg (Robust Adaptive Melt Pool Segmentation) algorithm, an intensity-agnostic and calibration-free edge-detection-based method that introduces three key innovations: (i) data-driven optimization of edge detection parameters, (ii) a quantitative segmentation review using an edge-to-area ratio to guide refinement, and (iii) an adaptive smoothing feedback loop. Unlike prior methods, RAMPSeg avoids arbitrarily selected parameters, over- or under-segmentation, and fixed heuristics. Instead, it achieves self-regulating segmentation that dynamically balances noise reduction and boundary preservation across diverse imaging conditions, enabling more generalizable melt pool analysis. We comprehensively evaluated its effectiveness using three distinct datasets encompassing diverse process conditions, materials, machines, and imaging systems. This unprecedented evaluation demonstrated consistently high segmentation accuracy (~90%) against expert-annotated ground truth, significantly outperforming conventional thresholding, Otsu's method, and state-of-the-art zero-shot vision transformer models (SAM and CLIPSeg). RAMPSeg enables reliable melt pool segmentation under dynamic conditions, supporting accurate anomaly detection, process optimization, and quality assurance.

29 - Poster - Topology-Aware Conformal Prediction for Stream Networks

Jifan Zhang, Northwestern University, Evanston, IL, United States

Stream networks, a unique class of spatiotemporal graphs, exhibit complex directional flow constraints and evolving dependencies, making uncertainty quantification a critical yet challenging task. Traditional conformal prediction methods struggle in this setting due to the need for joint predictions across multiple interdependent locations and the intricate spatio-temporal dependencies inherent in stream networks. Existing approaches either neglect dependencies, leading to overly conservative predictions, or rely solely on data-driven estimations, failing to capture the rich topological structure of the network. To address these challenges, we propose Spatio-Temporal Adaptive Conformal Inference (STACI), a novel framework that integrates network topology and temporal dynamics into the conformal prediction framework. STACI introduces a topology-aware nonconformity score that respects directional flow constraints and dynamically adjusts prediction sets to account for temporal distributional shifts. We provide theoretical guarantees on the validity of our approach and demonstrate its superior performance on both synthetic and real-world datasets. Our results show that STACI effectively balances prediction efficiency and coverage, outperforming existing conformal prediction methods for stream networks.

30 - Efficient Derivative Computation in Physics-Informed Neural Networks Without Autodiff

Soumik Sarkar, Florida State University, Tallahassee, FL, United States

no abstract

31 - Poster - Theoretical Analysis for Optimal Design of An Online Monitoring and Sampling Scheme Under Partial Observations**Dongmin Li, Georgia Institute of Technology, Atlanta, GA, United States**

High-dimensional streaming data is collected in a wide range of applications, such as advanced manufacturing and environmental monitoring. However, practical constraints such as limited sensors, transmission bandwidth, and data storage capacity, restrict the ability to observe all variables of interest at each data collection time for online monitoring. Although adaptive sampling strategies have been proposed that are integrated into monitoring schemes for quick anomaly detection, the theoretical analysis of detection performance for optimal design in practical settings has not been addressed. While few studies provide upper bounds for the average detection delay, the bounds are either asymptotic or excessively large, failing to provide valuable guidance on parameter design for optimal detection performance in practical applications. To fill this gap, we propose a holistic monitoring-based probabilistic sampling framework under partial observations and theoretically analyze the average detection delay to inform optimal design. By leveraging the sampling probabilities of variables, which are explicit functions of monitoring statistics, we derive formulas for practical average detection delay as functions of design parameters and system settings, such as the number of sampled variables and total variables in the system. These theoretical results enable the determination of the number of sampled variables and parameter settings to achieve optimal detection performance while satisfying false alarm requirements and considering practical resource constraints such as sampling costs. Simulation studies across various parameters and system settings validate the effectiveness of the theoretical results, demonstrating significant practical value for optimal performance design.

32 - Generative Conformal Prediction with Vectorized Non-Conformity Scores**Minxing Zheng, Carnegie Mellon University, Pittsburgh, PA, United States**

no abstract

33 - Heart: Heterogeneous Data-Driven Emotion and Anomaly Recognition in Sparse Longitudinal Texts**Prasanjit Dubey, Georgia Institute of Technology, Atlanta, GA, United States**

no abstract

34 - Autoregressive Neural Operator for Droplet Trajectory Prediction in Additive Manufacturing**Jinghao Cao, New Jersey Institute of Technology, Bayonne, NJ, United States**

no abstract

SD01

Building A Level 3 A301

George Nicholson Student Paper Competition

Award Session

Nicholson Student Paper Competition

Chair: Xinyun Chen, Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A

Co-Chair: Wolfram Wiesemann, Imperial College Business School, London, United Kingdom

1 - Reinforcement Learning in MDPs with Information-Ordered Policies**Zhongjun Zhang, Northwestern University, Evanston, IL, United States**

We introduce an epoch-based reinforcement learning algorithm for Markov Decision Processes in the infinite-horizon average cost setting leveraging a partial order over a policy class parameterized by $\in \Theta$. θ^* whenever data collected under can counterfactually estimate the performance of θ^* , without extra environment interactions. This yields a $(\sqrt{\log(|\Theta|)})$ regret where w is the order's width. We demonstrate our approach on several case studies in operations research, providing new theoretical guarantees and numerical validation.

2 - Robust Paths: Geometry and Computation**Hao Hao, Carnegie Mellon University, Pittsburgh, PA, United States**

Applying robust optimization requires calibrating the uncertainty set. In practice, this calibration is expensive and typically done by solving the robust counterpart multiple times under different uncertainty sets. We study the robust path: a set of robust solutions under varying the uncertainty set's parameters. Our central insight: the robust path can be algorithmically characterized by the trajectory of a proximal point method (PPM) applied to the nominal problem. Our results imply that a single run of an optimization algorithm on the nominal problem can approximate the entire robust path.

3 - A Stochastic Growth Model for Online Platform**Farbod Ekbatani, Chicago Booth School of Business, Chicago, IL, United States**

We investigate the growth trajectory of online platforms operating as market makers in two-sided matching markets. Their expansion, often referred to as the “chicken-and-egg dilemma”, poses an inherent challenge; service providers find value in joining the platform due to the presence of customers, while customers are drawn to the platform by the availability of diverse services. Under mild conditions, we characterize the optimal server compensation schemes the platform should adopt and show that the platform must account for each server's seniority, measured by the time at which they joined.

4 - Managing Tail Risk in Online Learning: Efficiency, Safety, Robustness, and Relationship to Alphago**Feng Zhu, MIT, Cambridge, MA, United States**

This work provides a detailed characterization of regret distribution in online learning, and demonstrates how managing regret tail risk leads to extra benefits on policy robustness. In Part I, we design policies that enjoy both optimal regret expectation and light-tailed regret distribution. In Part II, we study the optimal trade-off between expectation and tail risk for regret distribution. Notably, we find that our policy design coincides with what was adopted in AlphaGo MCTS, providing high-level insights to the success of their engineered solution in complex decision-making environments.

SD02

Building A Level 3 A302

Leveraging INFORMS Resources for Graduate Students

Panel Session

Committee's Choice

Co-Chair: Richard Staats, Retired at Retired, 11148 Byrd Dr, Mclean, VA, 22030, United States

1 - Panelist**Gokce Palak, Wake Forest University, Winston-Salem, NC, United States****2 - Panelist****Richard Staats, Retired at Retired, Mclean, VA, United States****3 - Panelist****Max Ji, The Pennsylvania State University, University Park, PA, United States****4 - Panelist****Jaime Mora-Vargas, Tec de Monterrey, Monterrey, Mexico**

SD03

Building A Level 3 A303

Multi-criteria Decision Aiding: Methods and Applications

Invited Session

Multi Criteria Decision Making

Chair: Milosz Kadzinski, Poznan University of Technology, Poznan, Poland

Co-Chair: Milosz Kadzinski, Poznan University of Technology, Poznan, Poland

1 - Multiple Criteria Decision Aiding Using a Rule Preference Model in Seven-Valued Logic**Roman Slowinski, Poznań University of Technology, Poznań, Poland, Salvatore Greco**

One of the challenges in Multiple Criteria Decision Aiding (MCDA) is ensuring the explainability and transparency of preference modeling. It is acknowledged that "if... then..." decision rules meet this requirement better than utility or outranking models. Moreover, rules induced from preference data using the Dominance-based Rough Set Approach (DRSA), when integrated with the recently proposed seven-valued logic, allow for distinguishing vagueness due to imprecision from that caused by inconsistency in the preference data. DRSA with seven-valued logic approach effectively addresses typical MCDA challenges, such as uncertainty, imprecision, and indeterminacy.

2 - Data Elicited from Pairwise Comparisons for Decision Making May Not be as Trustworthy as Believed**Evangelos Triantaphyllou, Louisiana State University, Baton Rouge, LA, United States, Juri Yanase**

A challenging issue in decision making is how to elicit qualitative / subjective data from decision makers. Using pairwise comparisons is a widely accepted approach as it offers an intuitive and appealing way towards this problem. The proliferation of this approach is mostly attributed to Saaty who proposed it as part of his AHP framework to decision making. Although many users may feel comfortable with this approach, a lingering question is whether one should rely on the derived results when pairwise comparisons are used. Thus, a critical question is how to validate it. We offer a testing approach that is based on a deliberately very optimistic (utopic) assumption. According to this assumption the decision maker is exceptionally accurate when he/she elicits the comparisons, one-at-a-time. We call it the ultra-accurate decision maker (or UADM) assumption. However, even under this highly favorable assumption towards the use of such comparisons, our research findings demonstrate that the results may be highly erroneous and thus it should not be trustworthy. The results may exhibit too often some serious logical contradictions.

3 - Triple-Loop Learning in Retail Demand Forecasting: a Systematic Evaluation of Benchmarking and Reproducibility of Machine Learning Models**Huijing Chen, University of Portsmouth, Portsmouth, United Kingdom, Huiqing Wang, Ashraf Labib**

In recent years, Machine Learning methods have gained more popularity in retail demand forecasting, aided by the availability of Big Data. Many algorithms have been proposed and applied, but there is no systematic understanding of how effective they are, and under what conditions. In this paper we incorporate the single-, double- and triple-loop learning theory as a theoretical lens in order to systematically categorise the retail demand forecasting literature from 2014 to 2024, especially on the goal of achieving triple-loop learning with three variant conceptualisations.

Two main issues identified from the literature are the lack of robust benchmarking and reproducibility. We apply MCDA tools to further analyse 60 papers on the two issues and provide insight and guidance for future research.

4 - Sensitivity Analysis for Multiple Criteria Decision Making: a New Perspective**Richard O'Shea, University College Cork, Cork, Ireland, Evangelos Triantaphyllou**

Multiple Criteria Decision Making (MCDM) methods such as the Weighted Sum Model (WSM), Compromise Programming (CP), and TOPSIS, combine the performance of alternatives across multiple criteria with criteria weights to generate a composite score and rank the alternatives. Sensitivity analysis is vital to assess the robustness of the ranking with regards to changes in criteria weights. Current sensitivity analysis methods use deterministic or stochastic alterations of weights and can be time consuming, computationally intensive, and imprecise. Existing analytical sensitivity analysis methods for the WSM can identify the most sensitive weight but only consider either the allowable increase or decrease, respectively. This work develops three analytical sensitivity analysis methods for use with the WSM, CP, and TOPSIS, respectively, to precisely determine the allowable increase and decrease to criteria weights. The most sensitive weight can be identified considering (i) the minimum allowable relative increase or decrease (ii) the weight stability interval (using the relative allowable increase and decrease) and (iii) a Pareto dominance approach considering relative increase and decrease. These methods are applied to a notional example. The allowable range of increase and decrease in criteria weights is larger when seeking to maintain the top q ranked alternatives than when seeking to maintain the full ranking of all alternatives. The most sensitive criteria weights can vary depending on the method used to define sensitivity and vary depending on whether the full ranking of alternatives is to be maintained, or, whether the ranking of the top q ranked alternatives is to be maintained.

5 - Integrating Response TIME and Attention Duration In Bayesian Preference Learning for Multiple Criteria Decision Aiding**Milosz Kadzinski, Poznan University of Technology, Poznan, Poland, jiaxuan jiang, Jiapeng Liu, Xiuwu Liao, Jingyu Dong**

We introduce a multiple criteria Bayesian preference learning framework incorporating behavioral cues for decision aiding. The framework integrates pairwise comparisons, response time, and attention duration to deepen insights into decision-making processes. The approach employs an additive value function model and utilizes a Bayesian framework to derive the posterior distribution of potential ranking models by defining the likelihood of observed preference data and specifying a prior on the preference structure. This distribution highlights each model's ability to reconstruct Decision-Makers' holistic pairwise comparisons. By leveraging both response time as a proxy for cognitive effort and alternative discriminability as well as attention duration as an indicator of criterion importance, the proposed model surpasses traditional methods by uncovering richer behavioral patterns. We report the results of a laboratory experiment on mobile phone contract selection involving real subjects using a dedicated application with time-, eye-, and mouse-tracking components. We validate the novel method's ability to reconstruct complete preferences. The detailed ablation studies reveal time- and attention-related behavioral patterns, confirming that integrating comprehensive data leads to developing models that better align with the DM's actual preferences.

6 - A Multi-Objective Optimization Approach to Optimize Dairy Farm Operations to Optimize Profit and Minimize Greenhouse Gas Emissions**Julio César Góez Gutiérrez, NHH Norwegian School of Economics, Bergen, Norway, Rasmus Bang, Mario Guajardo**

Governments and societies strive to balance food production with environmental goals. In this context, identifying farm-level economic and climatic win-win and lose-win scenarios is essential, not only for farmers but also for policymakers and the broader society. This study aims to examine economic performance and greenhouse gas (GHG) emission intensity in combined milk and beef production on dairy farms. We present a multi-objective optimization framework to economically optimize production at farm level under various operational constraints while minimizing the corresponding GHG emissions. We apply this framework in Norway. The optimization results demonstrate that the effectiveness of various measures to increase economic performance and to reduce GHG emissions strongly depends on operational conditions and the GHG emissions intensity measure used. We compare the results using GHG emissions intensity measure and analyze the impact on the managerial insights to support the dairy farmers.

SD04

Building A Level 3 A304

Blockchain Economics and Decentralized Finance

Invited Session

Finance

Chair: Fayçal Drissi, University of Oxford, Oxford, N/A, United Kingdom

1 - A Structural Model of Automated Market Making

David Cao, University of Pennsylvania, Philadelphia, PA, United States, Brett Falk, Leonid Kogan, Gerry Tsoukalas

Automated market makers (AMMs) facilitate ~\$1 trillion of transactions yearly, yet their reliance on static fees contrasts with classical microstructure theory, which emphasizes volatility-sensitive spreads in mitigating adverse selection. To assess whether this theory extends to AMMs, we develop a structural model capturing the interconnected decisions of noise traders, arbitrageurs, and liquidity providers. We empirically test this model using transaction-level data from Uniswap, leveraging a two-stage GMM approach to overcome endogeneity between trading and liquidity volumes. We find the optimal AMM fee structure dynamically adjusts to volatility, leading to improved trade. Our results suggest that, despite their microstructure differences, both traditional and decentralized markets can share fundamental design principles.

2 - Optimal Dex Convexity in the Presence of Infrastructure Rent: a Convex Analysis Approach

Nan Chen, The Chinese University of Hong Kong, Hong Kong, Hong Kong

Blockchain-based decentralized exchanges (DEX) are subject to infrastructure rents, imposing significant costs for liquidity provision. Utilizing convex analysis and calculus of variation methods, this work investigates how to select optimal convex pricing curves to mitigate adverse selection risks for DEXs.

3 - Do Longer Block Time Impair Market Efficiency in Decentralized Markets?

Fayçal Drissi, Oxford University, Oxford, United Kingdom, Alvaro Cartea, Agostino Capponi

In blockchains, traders bid priority fees to compete for queue position at regular intervals known as block time. Priority fees reduce trading volumes of informed traders and mitigate adverse selection costs. Longer block times can improve price efficiency by increasing speculative profits and by encouraging traders to acquire information. However, beyond a certain block time, expected losses from adverse selection outweigh the benefits of priority fees, leading to a liquidity freeze—unless uninformed liquidity demand grows sufficiently with block time. In equilibrium, traders only reveal their information immediately before the block is created.

4 - Designing Block Priority Mechanisms for Efficient Decentralized Markets

Gianluca Palmari, Scuola Normale Superiore, Pisa, Italy, Fayçal Drissi, Alvaro Cartea, Fabrizio Lillo

This paper investigates alternative designs for competition over block priority in blockchain systems and their implications for the microstructure of decentralised exchanges. We propose contest-based mechanisms in which higher priority fees increase the likelihood of obtaining priority in the block, but do not guarantee it. We develop a multi-stage model of trader competition and characterise equilibrium outcomes in terms of priority fees, trading volumes, and market participation. We compare the economic implications of different designs for builders, traders, and liquidity providers.

5 - Endogenous Price Formation at a Decentralized Exchange

Fahad Saleh, University of Florida, Winston-Salem, NC, United States

We provide an economic model of a Decentralized Exchange (DEX) liquidity pool to study price formation in this context. Crucially, we do not model an alternative venue (e.g., a Centralized Exchange) where the price formation occurs and thus our framework reveals the implications of price formation occurring directly at the Decentralized Exchange. We allow that the DEX employs a generalized Constant Function Automated Market Maker (CFAMM). Moreover, we allow for both strategic trading and endogenous liquidity provision. Our framework highlights the role that two-way positive externalities play in the price formation process.

SD05

Building A Level 3 A305

JFIG Paper Competition II

Award Session

Junior Faculty Interest Group

Chair: Karmel Shehadeh, University of Southern California, Los Angeles, CA, United States

Co-Chair: Raed Al Kontar, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Ozlem Ergun, Northeastern University, Newton, MA, United States

1 - Multi Agent Markov Entanglement

Shuze Chen, Columbia University, New York City, NY, United States, Tianyi Peng

Value decomposition has long been a fundamental technique in multi-agent dynamic programming and reinforcement learning (RL). Specifically, the value function of a global state $s_1, s_2, \dots, s_N$ is often approximated as the sum of local functions: $V(s_1, s_2, \dots, s_N) \approx \sum_{i=1}^N V_i(s_i)$. This approach traces back to the index policy in restless multi-armed bandit problems and has found various applications in modern RL systems. However, the theoretical justification for why this decomposition works so effectively remains underexplored.

In this paper, we uncover the underlying mathematical structure that enables value decomposition. We demonstrate that a multi-agent Markov decision process (MDP) permits value decomposition if and only if its transition matrix is not “entangled” — a concept analogous to quantum entanglement in quantum physics. Drawing inspiration from how physicists measure quantum entanglement, we introduce how to measure the “Markov entanglement” for multi-agent MDPs and show that this measure can be used to bound the decomposition error in general multi-agent MDPs. Using the concept of Markov entanglement, we prove that a widely used class of index policies is weakly entangled and enjoys a sublinear $\mathcal{O}(\sqrt{N})$ scale of decomposition error for N -agent systems. Finally, we show how Markov entanglement can be efficiently estimated in practice, providing practitioners with an empirical proxy for the quality of value decomposition.

2 - Learning to Select and Rank from Choice-Based Feedback: a Simple Nested Approach

Yifan Feng, National University of Singapore, Singapore, Singapore, Junwen Yang

We study a ranking and selection problem with choice-based feedback under dynamic assortments. In this setting, a company sequentially presents subsets of items to customers and observes their choices. Assuming only that preferences follow an unknown strict ranking, the goal is to identify the most preferred item or recover the full ranking using as few samples as possible and with high confidence. We propose two simple and effective algorithms: Nested Elimination (NE) for identifying the best item, and Nested Partition (NP), a divide-and-conquer extension of NE, for recovering the full ranking. Both algorithms come with sharp, instance-dependent, non-asymptotic sample complexity bounds. These bounds are derived using an analytical framework that characterizes the system dynamics as a sequence of multi-dimensional random walks. We also connect our nested approach to information-theoretic lower bounds. As a result, we show that NE is worst-case asymptotically optimal, while NP is near-optimal up to constant factors. Experiments on both synthetic and real-world data support our theoretical findings.

3 - Minimax Optimal Dynamic Assortment Selection with Contextual Information

Min-hwan Oh, Seoul National University, Gwanak-Gu, Korea, Republic of, Joongkyu Lee

We study the contextual multinomial logistic (MNL) bandit problem, where a decision-making agent sequentially selects assortments of items based on contextual information, and user choices follow an MNL model. This problem plays a central role in today’s online retail, online advertising, and recommendation systems, and is becoming increasingly relevant in the post-training of large language models. However, a long-standing open problem remains: (i) the gap between known lower and upper regret bounds, and (ii) the gap between theoretical guarantees and practical algorithms.

We close these gaps by establishing, for the first time, minimax optimal regret with matching lower and upper bounds. We design constant-time algorithms achieving the optimal regret, and we further provide instance-dependent optimal bounds. Our methods are both provably optimal (up to logarithmic factors) and computationally efficient. Empirical studies validate the theory, showing consistent improvements over existing approaches. To our knowledge, this is the first work to establish minimax optimality for contextual MNL bandits in both uniform and non-uniform revenue settings, while simultaneously providing practical algorithms.

4 - Online Learning for Equilibrium Pricing in Markets Under Incomplete

Devansh Jalota, Columbia University, Severna Park, MD, United States, Haoyuan Sun, Navid Azizan

The computation of equilibrium prices at which the supply of goods matches their demand typically relies on complete information on agents' private attributes, e.g., suppliers' cost functions, which are often unavailable in practice. Motivated by this practical consideration, we consider the problem of learning equilibrium prices over a horizon of T 's periods in the incomplete information setting wherein a market operator seeks to satisfy the customer demand for a commodity by purchasing it from competing suppliers with cost functions unknown to the operator. We first consider the setting when suppliers' cost functions are fixed and develop algorithms that, on three pertinent regret metrics, simultaneously achieve a regret of $\mathcal{O}(1)$ when the customer demand is constant over time, and $\mathcal{O}(\sqrt{T})$ when the demand varies over time. In the setting when the suppliers' cost functions vary over time, we demonstrate that, in general, no online algorithm can achieve sublinear regret on all three metrics. Thus, we consider an augmented setting wherein the operator has access to hints/contexts that reflect the variation in the cost functions and propose an algorithm with sublinear regret in this augmented setting. Finally, we present numerical experiments that validate our results and discuss various model extensions.

SD06

Building A Level 3 A306

Undergraduate Operations Research Prize

Award Session

Undergraduate Operations Research Prize

Chair: Ozge Yapar, Indiana University, Bloomington, IN, United States

1 - Prescriptive Machine Learning for Multi-Period Dynamic Nurse Deployment

Siqiang Wang, Peking University, Beijing, China, People's Republic of, Kailai Chen

To address staffing shortages, hospitals increasingly use float pool programs to deploy internal nurses dynamically across facilities. We formulate it as a multi-stage stochastic program and develop a prescriptive machine learning framework that combines offline optimization with real-time neural decision models. By embedding key structures, we ensure feasibility and solution quality. Trained on high-quality data, our method delivers near-optimal, constraint-satisfying decisions in milliseconds. A case study on real data demonstrates substantial speedups with minimal cost gaps.

2 - Survey of Data Driven Newsvendor: Unified Analysis and Spectrum of Achievable Regrets

Zhuoxin Chen, Tsinghua University, Beijing, China, People's Republic of

This paper studies the data-driven Newsvendor problem under all combinations of additive vs. multiplicative regret, high probability vs. expectation bounds, and different distribution classes. We provide a unified analysis based on a notion of clustered distributions, which shows that the entire spectrum of regrets between $1/\sqrt{n}$ and $1/n$ is possible. Our framework simplifies existing proofs and fills gaps in the literature. Simulations on commonly-used distributions demonstrate that our notion is the "correct" predictor of empirical regret across varying data sizes.

3 - An Aggregated Partition-Based Method for Decision-Focused Learning: Application to Microgrid Energy Management

Cristofer Lizama, Universidad Adolfo Ibáñez, Quilpué, Chile

This research addresses an integrated learning and optimization framework in data driven problems, in which an estimator is trained to achieve the best operational performance within a specific planning problem. The resulting model is formulated as a bilevel optimization problem with multiple lower-level subproblems. Given the high computational burden, an adaptation of the Aggregated Partition-Based Method is proposed to reduce the number of subproblems. The methodology is validated in a case study involving the energy management of residential microgrids with photovoltaic generation.

4 - Conformal Robust Optimization and Satisficing for Prescriptive Analytics

Lingjie Zhao, Tsinghua University, Haidian, China, People's Republic of

Prescriptive analytics increasingly leverages rich contextual data, yet classical predict-then-optimize fails under uncertainty. We introduce Conformal Robust Optimization (CRO) and Conformal Robust Satisficing (CRS), which learn and calibrate the parameter uncertainty sets with any, including black-box, predictor. We prove finite-sample coverage guarantees, analyze sample efficiency, derive suboptimality bounds for CRO, and establish a fundamental CRO–CRS equivalence. Numerical studies on knapsack and facility location problems show that CRO and CRS significantly outperform baselines.

SD07

Building A Level 3 A307

Dynamic and Robust Optimization in Transportation

Job Market Showcase

Job Market Showcase

Chair: Ahmad Ghasemkhani, University of Oklahoma, Norman, OK, 73071, United States

1 - Dynamic Crowdsourcing for Last-Mile Delivery with Parcel Lockers

Ali Sabzevari Zadeh, University of Alabama, Tuscaloosa, AL, United States, Iman Dayarian

This study addresses a dynamic crowdsourcing and last-mile delivery problem that integrates same-day delivery through parcel lockers (PLs) with next-day home delivery (HD), allowing PL customers to participate in delivery operations. At checkout, a PL customer selects a period from an available menu to schedule their pickup and may submit bids on bundles of HD orders from a customized list offered to them. If a bid is accepted, the associated bundle is delivered from the depot to the corresponding PL along with the customer's own orders, which they then deliver to the respective HD customers in exchange for store credits. Vehicle routes are periodically planned to jointly deliver selected HD orders to customers and PL orders, along with accepted HD bundles, to designated PLs before scheduled pickup times. Therefore, in each period, the retailer must decide which bids to accept, which HD orders to dispatch, and which to postpone, aiming to minimize delivery costs and credits awarded over a single-day, multi-period horizon. To solve the problem, which is stochastic due to the dynamic nature of PL order arrivals and the resulting variability in crowdsourcing capacity, a two-phase heuristic is developed. The first phase selects a subset of HD orders and bids using a sample-scenario planning approach to predict future PL orders and crowdsourcing capacity. The second phase designs vehicle routes for joint HD and PL delivery. Results show that cost savings decline as the HD-to-PL order ratio increases, especially in dense areas with sparse PL networks.

2 - Dynamic Path Planning in Stochastic Environments

Li Zhou, Auburn University, Auburn, AL, United States, Elvan Ceyhan

We introduce a sequential decision-making framework for navigation in environments with uncertain, spatially correlated obstacles. Using Gaussian Random Field models, our method captures spatial dependencies and continuously refines obstacle information through sequential Bayesian updates. By combining model-based and model-free reinforcement learning approaches, the proposed framework dynamically balances immediate traversal costs with the long-term benefits of information gathering, effectively addressing the exploration-exploitation tradeoff. Additionally, a search space reduction technique is integrated to enhance computational scalability.

Experimental evaluations demonstrate that our approach robustly adapts to dynamic uncertainties and presents improved performance across various simulation scenarios under different uncertainty, noisy conditions.

3 - Enhancing Adaptive Robust Optimization with Archetypal Analysis for Data-Driven Decision-Making

Ahmad Ghasemkhani, University of Oklahoma, Norman, OK, United States, Charles Nicholson, Andres Gonzalez Huertas

Robust optimization (RO) addresses decision-making under uncertainty by optimizing for worst-case scenarios within defined uncertainty sets. Archetypal analysis (AA), an unsupervised learning method that identifies extreme patterns in data, offers a powerful approach to construct these sets by simplifying complex datasets into representative scenarios. This abstract explores the integration of AA with RO to enhance decision-making in fields such as infrastructure restoration, economic resilience, and resource allocation. By applying AA, high-dimensional uncertain data such as infrastructure damage scenarios or economic disruption patterns are reduced to a small set of archetypes, each representing extreme yet plausible outcomes. These archetypes form the basis of uncertainty sets, enabling RO models to efficiently balance robustness and computational tractability. For instance, in interdependent network restoration, AA-derived scenarios capture critical failure patterns, guiding optimal resource allocation to minimize worst-case service disruptions. Similarly, in economic planning, archetypes highlight sector-specific resilience extremes, informing adaptive strategies. The synergy of AA and RO improves model interpretability and scalability, offering insights into uncertainty structures while maintaining rigorous optimization. This approach outperforms traditional uncertainty set constructions by leveraging data-driven insights, making it suitable for real-world applications with limited computational resources.

SD08

Building A Level 3 A310

AI, Algorithms, and Recommender Systems

Invited Session

Information Systems

Chair: Xiang (Shawn) Wan, Santa Clara University, Santa Clara, CA, 95050, United States

Co-Chair: Yinliang Tan, University of Houston, Houston, TX, United States

1 - Algorithmic Collusion or Competition: The Role of Platforms Recommender Systems

XINGCHEN XU, University of Washington, Seattle, WA, United States, Stephanie Lee, Yong Tan

Recent scholarly work has extensively examined the phenomenon of algorithmic collusion driven by AI-enabled pricing algorithms. However, online platforms commonly deploy recommender systems that influence how consumers discover and purchase products, thereby shaping the reward structures faced by pricing algorithms and ultimately affecting competition dynamics and equilibrium outcomes. To address this gap, we first develop a structural search model to characterize consumers' decision-making processes in response to varying recommendation sets. This model incorporates both observable and unobservable heterogeneity in utility and search cost functions, and is estimated using real-world data. Building on the resulting consumer model, we formulate personalized recommendation algorithms designed to maximize either platform revenue or consumer utility. We further introduce pricing algorithms for sellers and integrate all these elements to facilitate comprehensive numerical experiments. Our experimental findings reveal that a revenue-maximizing recommender system intensifies algorithmic collusion, whereas a utility-maximizing recommender system encourages more competitive pricing behavior among sellers. Intriguingly, and contrary to conventional insights from the industrial organization and choice modeling literature, increasing the size of recommendation sets under a utility-maximizing regime does not consistently enhance consumer utility. Moreover, the degree of horizontal differentiation moderates this phenomenon in unexpected ways. The "more is less" effect does not arise at low levels of differentiation, but becomes increasingly pronounced as horizontal differentiation increases. These findings provide valuable insights for regulators, platform designers, and market participants, and they open up a range of promising avenues for future research.

2 - Voice Your Preferences: the Dual Impact of Personal Voices on Personalized Recommendation

Yaqiong Wang, Santa Clara University, Santa Clara, CA, United States

As recommender systems evolve beyond traditional behavioral data to enhance personalization, understanding the broader impacts of using sensitive personal information becomes critical. This study examines the efficacy and implications of using personal voices in recommender systems. Through two controlled experiments, we first establish the significant correlation between individual vocal features and personal preferences. This correlation is mediated by individual demographic, emotional, and personality traits. These findings lay a theoretical foundation of recommender system design with personal voice beyond traditional behavioral data. Following the discovery, we investigate the impact of personalization strategy disclosure involving different types of personal data. Our results indicate that revealing the use of personal data—particularly voice—significantly reduces users' perceived recommendation quality and willingness-to-pay. This behavioral backlash highlights a growing tension between personalization and privacy, especially when biometric data is involved. Our findings offer empirical insights for system design and data regulation, emphasizing the significance of transparency strategies in the deployment of recommender systems.

3 - Decoding User Engagement: a Mutual Information-Enhanced Multi-Task Learning Framework for Predicting Gratification on Social Media

Cong Wang, Peking University, Beijing, China, People's Republic of, Xinpei Yao, Xunhua Guo

User engagement is a critical factor in the digital economy, where social media platforms strive to deliver contents that fulfill users' multifaceted gratification needs. These needs are inherently complex and linked to distinct engagement outcomes, making their accurate prediction essential for optimizing content delivery. Traditional multitask learning frameworks often struggle to model the intricate interdependencies among engagement objectives, particularly in data-sparse contexts with the absence of explicit relationships between these outcomes. This limitation can result in inferior multitask learning performance with "seesaw effect", where improving one objective inadvertently compromises another. To address these challenges, we propose the Mutual Information-Enhanced Multitask Graph Convolution Network (MIEMGCN), a novel approach for predicting multiple user engagement outcomes while capturing their complex interdependency. MIEMGCN integrates a multitask learning architecture with graph convolution networks, modeling both task-specific and shared embeddings to learn various engagement outcomes. This design allows MIEMGCN to effectively capture the social propagation of user preference and address data sparsity issues. Furthermore, MIEMGCN incorporates mutual information constraints into its loss function to enhance information sharing across tasks. These constraints quantify task correlations and promote more targeted and efficient information sharing. To overcome the computational challenges of exact mutual information inference, we propose a proxy based on conditional mutual information and provide theoretical validation of its effectiveness. Empirical evaluations on datasets from two major short video platforms demonstrate that MIEMGCN outperforms state-of-the-art methods by achieving higher prediction accuracy while mitigating the seesaw effect, offering valuable insights for boosting user engagement and refining content delivery strategies.

4 - When Product Recommendations Fall Short: the Heterogeneous Impact of Product Network Characteristics on Recommender System Performance

Xiang (Shawn) Wan, Santa Clara University, Santa Clara, CA, United States, Anuj Kumar, Haldun Aytug

Recommender systems create a network of interconnected products on e-commerce websites by showing related products on the focal product pages. We conduct a field study on a US fashion e-tailer's website to show how two core product network characteristics – the focal product's network centrality and the strengths of the edge connecting the focal product to its recommended products – could influence the recommendation's effect on product sales. While we find it beneficial to show either closely or weakly related recommended products for the focal products at the center of the recommendation network, it is helpful to recommend only weakly related products for the focal products at the edges. We provide empirical evidence that consumers seek a higher level of variety when exploring focal products at the network's edges. Thus, offering dissimilar product recommendations to them may better fulfill their variety-seeking needs. Finally, we show that retailers can exploit the heterogeneous impact of product network characteristics to generate recommendation policies to optimize product sales. We show the e-retailer could obtain four percent higher sales under our suggested recommendation policy.

5 - Personalization, Consumer Search, and Algorithmic Pricing

Liyang Qiu, Carnegie Mellon University, Pittsburgh, PA, United States, Yan Huang, Param Singh, Kannan Srinivasan

Our study investigates the impact of product ranking systems on artificial intelligence (AI) powered pricing algorithms. Specifically, we examine the effects of "personalized" and "unpersonalized" ranking systems on algorithmic pricing outcomes and consumer welfare. Our analysis reveals that personalized ranking systems, which rank products in decreasing order of consumer's utilities, may encourage higher prices charged by pricing algorithms, especially when consumers search for products sequentially on a third-party platform. This is because personalized ranking significantly reduces the ranking-mediated price elasticity of demand and thus incentivizes to lower prices. Conversely, unpersonalized ranking systems lead to significantly lower prices and greater consumer welfare. These findings suggest that even in the absence of price discrimination, personalization may not necessarily benefit consumers since pricing algorithms can undermine consumer welfare through higher prices. Thus, our study highlights the crucial role of ranking systems in shaping algorithmic pricing behaviors and consumer welfare.

SD09

Building A Level 3 A311

Novel Approaches to Data Analysis

Job Market Showcase

Job Market Showcase

Chair: Ann-Kathrin Meyer, University of Münster, Münster, Germany

1 - Dynamic Correlation Learning in Diverse Markets: a Graph-Based Approach for Adaptive Stock Price Prediction

Fang He, Southeast University, Jiangsu, China, People's Republic of, Wei Yin

Stock price forecasting has become a significant and complex research area within financial technology. The dynamic correlations among stocks and the inherent noise in price volatility present considerable challenges in accurately forecasting stock prices and enhancing investment returns. This paper introduces a novel Dynamic Correlation Graph Convolution Network (DyCGCN) with embedded temporal correlation extraction. Firstly, we propose a dual-scale dynamic graph generation method to capture the topological relationships among stocks. Secondly, we develop a dynamic correlation-temporal convolution module that extracts high-level temporal correlations. Thirdly, we introduce a prospect theory-guided multi-strategy loss function that accommodates the diverse risk preferences of investors. Additionally, we present a joint regression-classification learning method to extract and leverage stock trend information. Experimental results on four real-world datasets show the efficacy of the proposed DyCGCN model, demonstrating its superior accuracy and effectiveness in stock price forecasting.

2 - Fairness, Balance, and Scale: a Multilevel Graph Approach to Deep Neural Networks

Rachel Bennett, University of Oklahoma, Oklahoma City, OK, United States

The rapid growth in data generation has elevated neural networks as essential tools in machine learning. However, this growth also brings challenges, including increased computational demands, class imbalance, and bias toward certain subpopulations. To address these issues, we propose a novel **Multilevel Deep Neural Network (MLDNN)** framework designed to scale efficiently to large datasets while addressing class imbalance. Our method uses graph-based data representations and coarsening techniques to progressively reduce the dataset size. To further enhance model fairness, we introduce a fairness-aware extension of this framework. This approach integrates fairness considerations during the coarsening stage to reduce bias across demographic groups. We evaluate our MLDNN models against conventional single-level deep neural networks on both benchmark and real-world datasets.

3 - Cracking the Ca(U)Ses: Unraveling Urban Crime Dynamics with Agent-Based Modeling

Ann-Kathrin Meyer, University of Münster, Münster, Germany, Konstantina Valogianni, Balaji Padmanabhan, Tobias Brandt

The complexity of phenomena like infectious diseases, traffic congestion, or crime arises from their intricate and diverse underlying relationships. While statistical models are powerful aids towards understanding those phenomena, they often fall short with respect to flexibility and addressing deeper causal structures. This issue is particularly relevant in

cases where different combinations of causal factors can lead to similar outcomes (equifinality) or similar initial conditions to different outcomes (multifinality). Agent-Based Modeling offers a promising alternative by employing a dynamic, process-oriented approach to causation. This is a simulation method that models the interactions of autonomous agents

to assess their effects on the system as a whole. Causal Agent-Based Modeling, i.e. grounded in theoretical and empirical foundations, facilitates a comprehensive understanding of diverse pathways leading to similar or different outcomes. Using the Causal Agent-Based Modeling lens, this paper addresses the complexities of modeling crime in an urban environment by incorporating spatial factors together with temporal dimensions. In addition to studying a key problem through a new lens, the exploration of equifinality and multifinality using Causal Agent-Based Modeling contributes to operations management research by accounting for the numerous interacting factors and uncertainties inherent in operational processes and policy decisions. We show that our approach can uncover multiple plausible causal sets for criminal behavior using real spatial data, rather than seeking a single explanation. This methodology therefore can aid investigators by offering automated candidate explanations, improving current crime intelligence practices and enhance operational decisions through explicit industry-specific contextual understandings.

SD10

Building A Level 3 A312

Modeling the Electricity Grid

Job Market Showcase

Job Market Showcase

Chair: Yamil Essus, North Carolina State University, Raleigh, NC, United States

1 - Centralized Scheduling of Home Charging Operations for Electric Vehicles: a Gamification Approach

Alireza Ghahtarani, HEC Montreal, Côte-St-Luc, QC, Canada, Jorge Mendoza, Martin Cousineau, Amir-massoud Farahmand

We study the centralized scheduling of home charging operations for electric vehicles (EVs) using a gamification-based approach. Inspired by Tetris, we model the charging schedule as a dynamic block-stacking game, where scheduling decisions impact grid efficiency. We propose heuristic methods and apply DAGGER (Dataset Aggregation) to train an agent that learns to optimize EV charging schedules through imitation learning. The primary objective is peak shaving, ensuring a balanced distribution of charging demand to reduce grid stress. Our approach is compared against reoptimization methods and full information optimization, evaluating trade-offs between computational efficiency and performance. Simulation results demonstrate that gamification offers a practical and scalable alternative, effectively mitigating peak loads while maintaining scheduling efficiency.

2 - Risk-Averse Markov Decision Processes for Managing Uncertainty in the Electricity Grid

Arash Khojaste, University of Massachusetts Amherst, Amherst, MA, United States, Jonathan Pearce, Golbon Zakeri

We present various methods from the perspective of system operators to enhance decision-making in an electricity grid facing the uncertainties associated with renewable energy sources. We incorporate different risk measures into our decision framework to address risk aversion in the system's operational planning. In a Markov Decision Process (MDP) framework, we use Conditional Value-at-Risk (CVaR) to build a risk-averse model that minimizes costs while ensuring grid reliability by hedging against high-cost scenarios. We propose a computationally efficient method to solve the resulting optimization problem and reach a global optimum. Additionally, we suggest various constraints that can increase the electricity grid's reliability. These constraints include one that prevents consecutive demand curtailments. We use a real-world example of a backup system for wind power integration to show the application of our models.

3 - Efficient Integration of Energy Storage Resources in Modern Power Systems Through Synergistic Integration of Mathematical Optimization and Machine Learning

Farhan Hyder, Rochester Institute of Technology, Rochester, NY, United States

The recent emergence and increased penetration of variable renewable energy resources introduce new operational challenges in the modern electrical grid. For the grid operators, this means more generation-demand imbalances, greater pressure on the transmission network, and increased computational burden for efficient operation of the grid. Utility-scale Energy Storage Resources (ESRs) can play a crucial role in mitigating the load fluctuations caused by renewables, as they are capable of instantly switching between charging (consuming) and discharging (generating) modes for a reliable, secure and sustainable grid of the future. However, with the integration of a large number of ESRs, the complexity of the resulting unit commitment (UC) problem increases significantly due to the increased number of binary variables for preventing simultaneous charging and discharging. Hence, computationally efficient modeling of ESRs can play a key role in mitigating the computational challenges introduced by the integration of ESRs.

To address the challenge of increased computational burden, a machine learning enhanced formulation tightening approach is developed. Under this approach, the ESR's mathematical formulation is transformed to be closer to its convex hull, resulting in better computational performance. However, due to the combination of utility-scale renewables and ESRs, the computational burden might still be high. Hence, a novel deep learning-based approach is developed to filter redundant transmission capacity constraints in the UC problem. This can improve computational performance of the UC problem.

Results on medium and large-sized networks shows the effectiveness of the formulation tightening and transmission line screening approaches.

4 - Leveraging GeoAI Foundation Models for Generalizable Spatial Prediction of Power Outages

Yamil Essus, North Carolina State University, Raleigh, NC, United States, Benjamin Rachunok

Uncertainty in the location of weather-based power outages leads to multidimensional risks associated with cascading impacts on engineering systems. The main challenge in developing generalizable models for power outage prediction is the high dimensionality of weather features representing multiple geographies. We applied transfer learning and fine-tuning to build a deep learning model based on recent Geospatial Artificial Intelligence (GeoAI) methods to produce accurate estimations of power outage locations. In order to reduce data and compute requirements, we used Prithvi WxC, a foundation model for weather and climate applications, as the starting point of our network.

SD11

Building A Level 3 A313

Inference and Resource Allocation in Healthcare Systems

Job Market Showcase

Job Market Showcase

Chair: Hyunwoo Shin, Virginia Tech, Blacksburg, VA, United States

1 - Optimal use of Alternative Care Modalities and Ppe Supplies During Infectious Disease Outbreaks

Mina Bahadori, Clemson University, Central, SC, United States, Tugce Isik

Resource shortages—whether of healthcare personnel or medical equipment—are a persistent challenge in healthcare systems. These shortages are often exacerbated during infectious disease outbreaks due to sudden surges in demand and supply chain disruptions, and tend to affect different facilities and patient populations unequally. In this work, we focus on the effective and fair use of limited personal protective equipment (PPE) supplies, which is critical to protect healthcare workers and maintain safe and successful care delivery. We formulate a finite-horizon Markov decision process (MDP) to optimally allocate different patient populations to alternative care modalities during a PPE shortage induced by an infectious disease outbreak. We consider in-person care with PPE and telehealth as available care modalities and maximize total utility of PPE allocation decisions across different patient populations who differ in infection risk and in their ability to receive care in-person or via telehealth. We analyze the structure of the optimal policy and establish key monotonicity properties of the reward and value functions. These structural insights enable us to design interpretable and implementable dynamic allocation policies. In addition, we explore how various factors—including inventory replenishment policy used by the healthcare facility and uncertainty in PPE replenishment (e.g., delays, partial restocking) impact the optimal policies. Our findings highlight the need for adaptive, risk-aware strategies to manage resource shortages and offer practical guidance for equitable decision-making during public health crises.

2 - Heterogeneous Treatment Effects in Panel Data: Insights into the Healthy Incentives Program

Emily Zhang, Massachusetts Institute of Technology, Cambridge, MA, United States, Retsef Levi, Elisabeth Paulson, Georgia Perakis

We study the impact of adding new vendors to the Massachusetts (MA) Healthy Incentives Program (HIP), a food-subsidy program aimed at increasing the consumption of fruits and vegetables among low-income households, on the utilization of the program. This is an instance of a core problem in causal inference: estimating heterogeneous treatment effects (HTE) using panel data. Existing methods either do not utilize the underlying structure in the panel data or do not accommodate the general vendor addition patterns in the HIP data. We propose the Panel Clustering Estimator (PaCE), a novel method that first partitions observations into disjoint clusters with similar treatment effects using a regression tree, and then leverages the (assumed) low-rank structure of the panel data to estimate the average treatment effect (ATE) for each cluster. Our theoretical results establish the convergence of the resulting estimates to the true treatment effects. Computational experiments with semi-synthetic data show that PaCE achieves superior accuracy for ATE and HTE estimation compared to existing approaches. This performance was achieved using a regression tree with no more than 40 leaves, making PaCE both more accurate and interpretable. Applying PaCE to HIP data, we discern the heterogeneous impacts of vendor additions on program utilization across different regions of MA. These findings provide valuable insights for future budget planning and for identifying which MA ZIP codes to target with vendor additions.

3 - Optimizing U.S. Lung Allocation Policy via a Multi-Objective Trust-Region-Based Simulation Optimization Algorithm

Hyunwoo Shin, Virginia Tech, Blacksburg, VA, United States, Sait Tunc, Xi Chen

Simulation models play a critical role in evaluating and guiding organ allocation policies. A common approach by using kriging-based surrogates with active learning faces two key challenges: difficulty in setting credible input bounds and the high computational burden of global exploration and repeated surrogate updates. To address these challenges, we propose a multi-objective trust-region simulation-optimization framework, tailored for lung allocation policy design. Our method begins each search iteration with a low-fidelity kriging surrogate, built from existing simulation data, to identify a promising initial policy. The search is then confined to a dynamically adjusted trust region around the current best solution, allowing focused exploration while limiting unnecessary simulations. Our localized strategy significantly reduces computational costs by avoiding repeated global surrogate fits and by targeting high-potential areas for refinement. This makes it a practical and scalable tool for policy optimization in real-world transplantation settings, supporting timely and evidence-based decision-making.

4 - Mass Vaccination Scheduling: Trading off Infections, Throughput, and Overtime

Shanshan Luo, UBC Sauder School of Business, Vancouver, BC, Canada

Mass vaccination is essential for epidemic control, but long queues can increase infection risk. We study how to schedule arrivals at a mass vaccination center to minimize a tri-objective function of a) the expected number of infections acquired while waiting, b) throughput, and c) overtime. Leveraging multimodularity results of a related optimization problem, we construct a solution algorithm and apply it to a case study of COVID-19. We find that although the standard equally-distributed, equally-spaced schedule sits near the Pareto-optimal frontier, it is located away from a sharp elbow in the trade-off between infections and overtime. Specifically, the “elbow policy” achieves approximately 38% fewer expected infections for nearly the same expected overtime. We also discuss managerial insights around the structure of the optimal schedule and compare it to the well-known “dome-shaped” policies found in other appointment scheduling settings.

SD12

Building A Level 3 A314

How to Make Decisions Under Uncertainty in the AI Era?

Panel Session

Applied Probability Society

Co-Chair: David Goldberg, Cornell University, Ithaca, NY, United States

1 - Moderator Panelist**David Goldberg, Cornell University, Ithaca, NY, United States**

This panel will discuss a variety of aspects arising from the above OR-AI comparison/conflict/interface, in how to make decisions under uncertainty, a question at the core of the entire OR discipline.

These include, for example: 1) What are the roles and strong/weak scenarios in using mathematical analysis, e.g., asymptotic analysis for large-scale stochastic systems, compared to using heuristic or black-box AI approaches in decision learning? 2) The usual "mindset" of OR methodological study is to devise algorithms by theoretically understanding their performances, through convergence analysis, regret or error bounds, computational complexity, etc. Would it be a competitive approach to empirically test and validate performances only or before asking for a rigorous theory? 3) Should, or how much should, we change the criteria to publish in top OR journals to embrace more empirical papers or judge the merit of theoretical works? 4) How do we compare the interpretability and trustworthiness of OR models with AI models, which may be opaque but empirically performant? Especially when facing high-stakes decisions, how should we trade off explainability vs. accuracy? 5) How should we train the next generation of OR researchers and practitioners in balancing theory, data skills, and domain knowledge needed to thrive at the OR-AI interface?

2 - Panelist**Devavrat Shah, Massachusetts Institute of Technology, Cambridge, MA, United States****3 - Moderator Panelist****4 - Panelist****Jing Dong, Columbia University, New York, NY, United States****5 - Panelist****Peter Frazier, Cornell University, Ithaca, NY, United States****6 - Panelist****Amy Ward, The University of Chicago Booth School of Business, Chicago, IL, United States****SD13**

Building A Level 3 A315

Learning, Ranking, and Selection

Job Market Showcase

Job Market Showcase

Chair: Zixian Yang, University of Michigan, Ann Arbor, MI, 48109, United States

1 - Bayesian Procedures for Ranking and Selection with Multiple Objectives**Jinbo Zhao, Texas A&M University, College Station, TX, United States**

Ranking and selection (R&S) problems with multiple objectives remain challenging when approached from a Bayesian perspective, primarily due to the computational difficulty of evaluating integrals over high-dimensional, non-convex regions. We propose a reformulation using sample average approximation via posterior sampling that effectively circumvents these complex integrals. The resulting framework supports diverse acceptability definitions—including top-m selection, stochastically constrained optimality, and Pareto optimality—provided that acceptability can be determined from the known performances of all simulated alternatives. We study several Bayesian guarantees concerning the probability of returning at least one acceptable alternative, the probability of returning a superset of all acceptable alternatives, and the expected proportion of acceptable alternatives returned. We further propose two adaptive budget allocation schemes: one based on the gradient of these posterior quantities and another based on the gradient of the largest posterior variance among alternatives' probabilities of being acceptable. Both gradients are with respect to additional simulation effort and estimated via posterior sampling. Preliminary numerical experiments illustrate the potential effectiveness and efficiency of these adaptive allocation policies in multi-objective R&S.

2 - Optimizing Input Data Collection for Ranking and Selection**Taeho Kim, HKUST, Hong Kong, China, People's Republic of, Eunhye Song**

We study a ranking and selection (R&S) problem when all solutions share common parametric Bayesian input models updated with the data collected from multiple independent data-generating sources. Our objective is to identify the best system by designing a sequential sampling algorithm that collects input and simulation data given a budget. We adopt the most probable best (MPB) as the estimator of the optimum and show that its posterior probability of optimality converges to one at an exponential rate as the sampling budget increases. Assuming that the input parameters belong to a finite set, we characterize the ϵ -optimal static sampling ratios for input and simulation data that maximize the convergence rate. Using these ratios as guidance, we propose the optimal sampling algorithm for R&S (OSAR) that achieves the ϵ -optimal ratios almost surely in the limit. We further extend OSAR by adopting the kernel ridge regression to improve the simulation output mean prediction. This not only improves OSAR's finite-sample performance, but also lets us tackle the case where the input parameters lie in a continuous space with a strong consistency guarantee for finding the optimum. We numerically demonstrate that OSAR outperforms a state-of-the-art competitor.

3 - Optimizing Adaptive Experiments: A Unified Approach to Regret Minimization and Best-Arm Identification**Chao Qin, Stanford University, Palo Alto, CA, United States, Daniel Russo**

Practitioners conducting adaptive experiments often encounter two competing priorities: maximizing total welfare (or 'reward') through effective treatment assignment and swiftly concluding experiments to implement population-wide treatments. Current literature addresses these priorities separately, with regret minimization studies focusing on the former and best-arm identification research on the latter. This paper bridges this divide by proposing a unified model that simultaneously accounts for within-experiment performance and post-experiment outcomes. We provide a sharp theory of optimal performance in large populations that not only unifies canonical results in the literature but also uncovers novel insights. Our theory reveals that familiar algorithms, such as the recently proposed top-two Thompson sampling algorithm, can optimize a broad class of objectives if a single scalar parameter is appropriately adjusted. In addition, we demonstrate that substantial reductions in experiment duration can often be achieved with minimal impact on both within-experiment and post-experiment regret.

4 - Online Learning for Two-Sided Markets**Zixian Yang, The University of Michigan, Ann Arbor, MI, United States, Sushil Mahavir Varma, Lei Ying**

We study a two-sided market, wherein, price-sensitive heterogeneous customers and servers arrive and join their respective queues. A compatible customer-server pair can then be matched by the platform, at which point, they leave the system. Our objective is to design pricing and matching algorithms that maximize the platform's profit, while maintaining reasonable queue lengths. As the demand and supply curves governing the price-dependent arrival rates may not be known in practice, we design a novel online-learning-based pricing policy and establish its near-optimality. In particular, we prove a tradeoff among three performance metrics: $\tilde{O}(T^{\frac{1}{2}})$ regret, $\tilde{O}(T^{\frac{1}{2}})$ average queue length, and $O(T^\gamma)$ maximum queue length for $0 < \gamma \leq 1/6$, where T is the time horizon. Moreover, barring the permissible range of γ , we show that this trade-off between regret and average queue length is optimal up to logarithmic factors under a class of policies, matching the optimal one as in (Varma et al., 2023), which assumes the demand and supply curves to be known. Our proposed policy has two noteworthy features: a dynamic component that

optimizes the tradeoff between low regret and small queue lengths; and a probabilistic component that resolves the tension between obtaining useful samples for fast learning and maintaining small queue lengths.

SD14

Building A Level 3 A316

Novel Problems in Revenue Management

Job Market Showcase

Job Market Showcase

Chair: Jessie Jiang, Cornell University, Ithaca, NY, United States

1 - Pricing with Tips in Three-Sided Delivery Platforms

Gary Ma, Harvard University, Cambridge, MA, United States, Yannai Gonczarowski, David Parkes

We model a delivery platform facilitating transactions among three sides: buyers, stores, and drivers. In addition to buyers paying store-specific purchase prices and drivers receiving store-buyer-specific delivery compensation from the platform, each buyer has the option to directly tip for delivery from a specific store. An equilibrium consists of prices, compensations, tips, and transactions that clears the market, such that buyers receive deliveries from preferred stores considering prices and tips they pay, and drivers deliver preferred orders considering the compensations and tips they receive.

We illustrate the role of tips in pricing: with tips an equilibrium always exists, but without tips an equilibrium is only guaranteed to exist when drivers outnumber the minimum number of buyers and stores. From an efficiency perspective, the optimal with-tip equilibrium welfare is always weakly larger than the optimal without-tip equilibrium welfare. However, we show that even with tips, efficient equilibria may not exist, and that calculating the optimal equilibrium welfare is NP-hard. To address these challenges, we identify conditions on market structure that ensure the existence of efficient with-tip equilibria and allow these efficient equilibria to be computed in polynomial time.

2 - Fluid Approximations in Revenue Management with Correlated Demands and Intertemporal Constraints

WEIYUAN LI, Cornell Tech, New York, NY, United States, Huseyin Topaloglu, Paat Rusmevichientong, Jingwei Zhang

In this talk, we explore fluid approximations as a tool for tackling revenue management problems with correlated demand and intertemporal constraints. We focus on designing asymptotically optimal policies informed by these fluid approximations.

3 - Beyond One-Size-Fits-All: Personalized Delivery and Fulfillment Optimization

Quan Zhou, McGill University, Montreal, QC, Canada, Mehmet Gumus, Sentao Miao

Motivated by our collaboration with an online platform operating in North America, we explore the joint optimization of the order fulfillment process with personalized delivery options in the context of e-commerce. Customers can choose from personalized fulfillment options to proceed with the purchase or leave with no purchase. The retailer periodically makes fulfillment decisions and relies on multiple logistic providers to perform the fulfillment operations. We model customer behavior with a general discrete choice model and formulate the joint optimization as a stochastic dynamic program. We propose a tractable deterministic approximation and develop a computationally efficient heuristic with a provable performance guarantee. Using real datasets collected from our industrial partner, we demonstrate the value of personalizing fulfillment options for the customers and jointly optimizing the options with fulfillment assignments. Our results show that demand management via personalized fulfillment options is prominent when customers favor quicker fulfillment and when the fulfillment capacity is limited. However, an optimized fulfillment operation becomes more critical when customers are more willing to wait.

SD15

Building B Level 3 B312

The One Where Operations Fights Climate Change

Invited Session

MSOM: Sustainable Operations

Chair: Siddharth Singh, University College London, London, E14 5AA, United Kingdom

1 - Preventing Catastrophic Wildfires Through Optimized Grid Annual Inspection and Maintenance Plan

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

Power grid failures are a major cause of catastrophic wildfires, making wildfire prevention a top priority for utility companies and regulators. Inspection and maintenance are crucial to wildfire mitigation, and utilities must annually report which grid segments will be inspected, the planned preventive actions, and anticipated reductions in wildfire costs. This paper addresses the inspection and maintenance planning problem to minimize wildfire costs due to grid failures. The power grid consists of millions of components grouped into thousands of clusters, each containing hundreds of correlated segments. Using uncertain wildfire cost estimates provided by utilities, we formulate a large-scale stochastic optimization problem to design optimal inspection and maintenance plans under capacity constraints. We show that the problem is NP-hard and analyze how inspection and segment correlation affect cost reduction. We reformulate it as a more tractable, non-convex integer program, which decomposes by cluster, and develop an approximate solution method that constructs a surrogate objective via the lower surface of the convex hull of sampled points. We also derive a lower bound on the objective and provide a theoretical a priori error analysis. We validate our framework using a disguised instance based on data from one of California's largest utility companies. Our method achieves an approximation error of just 0.22%, significantly outperforming a heuristic similar to current industry practice, which incurs a 14.31% error.

2 - Effect of Charging Stations on EV Adoption

Andrey Makhnov, University of North Carolina, Chapel Hill, NC, United States, Nur Sunar, Jayashankar Swaminathan

The strategic deployment of electric vehicle (EV) charging infrastructure plays a critical role in adoption, yet optimizing placement remains a complex challenge. This research applies a state-of-the-art machine learning approach to evaluate the impact of charger locations and network density on adoption patterns. By leveraging geospatial and socioeconomic modeling, we estimate heterogeneous treatment effects of charging station deployment across different regions. Findings from this analysis aim to provide insights that support data-driven decision-making for infrastructure optimization.

3 - Subsidizing Circularity to Support the Domestic Renewable Energy Industry

Nilsu Uzunlar, Carnegie Mellon University, Pittsburgh, PA, United States, Alan Scheller-Wolf, Siddharth Singh

There is expected to be a large number of solar panels and wind turbines disposed of in the upcoming decade. These products are prohibitively expensive to recycle. In addition to the problem of expensive recycling, some countries' abundant production/mining capacity makes domestic manufacturers overly reliant on foreign suppliers. We explore how domestic recycling regulation could solve the dual problems of catalyzing domestic production and reducing renewable energy product waste. We propose a model in which a regulator sets a recycling requirement for key commodities. Simultaneously, the regulator introduces competition by subsidizing a domestic recycler, who operates alongside a foreign supplier. We use a game-theoretic framework to understand the supplier's capacity decision in response to the introduction of the recycler, how this affects demand allocation within the market, and its implications for renewable energy supply and mining. We fully characterize the optimal recycling requirement for a focal commodity. We find that it is infeasible to impose legislation if the collection rate of retiring products is too low, or if the expected price of the commodity is too high. In general, as the expected market price increases, both the collection rate and the subsidy must increase to render recycling effective.

Furthermore, an alternative intervention, imposing tariffs on foreign commodity supply, will typically not render recycling feasible if it was infeasible to start with. However, if legislation is feasible without a tariff, it may remain so under a tariff, and the tariff may potentially lead to an increase in recycling volumes.

4 - Watt's Next: Navigating the Surge in Electric Vehicle Adoption and Its Implications for Grid Operations

Siddharth Singh, University College London, London, United Kingdom, Karthik Murali, Owen Wu, Mesut Yavuz

As policy decisions pave the way for more Electric Vehicles (EVs), a tremendous strain on the utility grid is expected. Managing this strain requires careful planning and possibly control of EV charging. This may include utilities managing intra-day demand using grid-scale batteries and influencing EV charging through active or passive managed charging programs, including possible use of Vehicle-to-Grid (V2G) technologies. Managed charging must be carefully controlled, accounting for EV users' driving patterns and the extent to which they permit the utility to manage their battery ("offered flexibility"). How can this best be done? We explore this question from the perspective of an integrated energy system planner seeking to minimize generation and emissions costs from the electricity and transportation sectors. Using a stylized analytical model with two daily periods (peak and off-peak), we derive optimal grid operations assuming a fraction of EVs participate in managed charging. We find the interplay between EVs and grid-scale batteries is non-trivial—they can be substitutes or complements in the planner's portfolio. We supplement our analytical model with a higher-fidelity 24-hour model calibrated to real data from the PJM interconnection region and extend it numerically to include EV drivers with heterogeneous default charging behavior. The most effective lever to reduce system costs is managing more EVs, rather than increasing flexibility of already-managed ones. With sufficient EV management, reliance on grid-scale batteries decreases, and the need for V2G—associated with high battery degradation—is avoided.

SD15

Building A Level 4 A401

Innovative Approach to Efficient Airspace Management

Invited Session

Air Transportation Section

Chair: Go Nam Lui, Lancaster University Management School, Lancaster, LA15NL, United Kingdom

1 - 3D Optimal Airspace Sector Design with Mixed Integer Linear Programming

Go Nam Lui, Lancaster University, Lancaster, United Kingdom, Guglielmo Lulli

The increasing complexity of air traffic management and the growing demand for efficient airspace utilization have driven the development of smart upper airspace systems that leverage advanced operations research methodologies. This presentation presents the application of operations research techniques in the design and configuration of intelligent upper airspace frameworks in the SMARTS project, focusing on optimization models that enhance traffic flow management and improve overall system performance. We explore two key applications: First, the optimal design of airspace sectors. Mathematical programming models, such as Mixed Integer Programming (MIP), are employed to create sectors that balance controller workload and ensure smooth traffic flow, with the connectivity and traffic flow convexity consideration. Complementary heuristic approaches demonstrate the ability to achieve near-optimal solutions rapidly, crucial for operational agility. Second, dynamic airspace configuration. Here, OR techniques, including integer linear programming and graph-based algorithms, enable the selection of optimal airspace configurations in response to fluctuating traffic demand and uncertainty. These methods aim to maximize capacity, minimize delays, and provide robust plans adaptable to unforeseen demand surges. Validated with real-world data, these OR applications provide theoretically sound and practically implementable tools. They offer significant improvements in ATM efficiency, workload distribution, and system robustness, paving the way for more adaptive and resilient airspace management in the face of evolving challenges.

2 - Functional Data Analysis in Flight Trajectory Clustering

GUGLIELMO LULLI, University of Milano-Bicocca, Milano, Italy, Nicos Pavlidis, Go Nam Lui, Sameer Alam, B Pang, C Raibulet

Machine learning (ML) is revolutionizing flight trajectory analysis, offering unprecedented capabilities to enhance safety, efficiency, and sustainability in aviation. This presentation showcases diverse ML applications that transform raw flight data into actionable intelligence. We explore how predictive models, leveraging techniques like regression and time-series analysis, can accurately forecast arrival times, fuel burn, and potential conflicts, enabling proactive decision-making. Unsupervised learning algorithms, such as clustering and anomaly detection, are examined for their power in identifying common flight patterns, detecting deviations from standard procedures, and flagging potential safety hazards from large-scale trajectory datasets (e.g., ADS-B). The presentation will demonstrate tangible benefits, including optimized route planning for reduced fuel consumption and emissions, improved airspace capacity utilization, and enhanced situational awareness for air traffic controllers and pilots. We will also touch upon the challenges of data integration and model validation in this critical domain. Ultimately, this exploration aims to underscore how ML-driven trajectory analysis contributes directly to a safer, more efficient, and environmentally conscious global air transportation system, paving the way for next-generation air traffic management. This work has been supported by the AI4ATM project, funded by both MAECI (Italy) and A*STAR (Singapore).

3 - Evaluation of near-Terminal Operation Performance Under Uncertainties in High-Density Airspace

Yutian Pang, The University of Texas at Austin, Austin, TX, United States

Future high-density airspace operations are sensitive to uncertainties from various aspects of the aviation system. This challenge is expected to intensify with the integration of highly automated and autonomous aircraft into the existing airspace. To address this, the present study evaluates near-terminal arrival performance under two sources of aeronautical communication uncertainty, (i) communication availability, (ii) communication system latency, as well as the uncertain human-induced pilot response delays. The analysis considers a congested arrival scenario with two converging flows one straight-in approach stream and one downwind stream merging via a base-to-final turn. Aircraft arrivals are generated by modified Poisson processes subject to minimum separation constraints, simulating high-density traffic operations. The merging logic incorporates fixed-duration turns, pilot response delays, and dynamic gap availability in real time. System performance is measured using runway throughput, the average delay (holding time) experienced by downwind aircraft, and the fraction of downwind arrivals that merge immediately without holding. The evaluation employs a Monte Carlo-based discrete event simulation to analyze system behavior across a wide range of conditions. To capture the impact of all scenarios, importance sampling techniques are incorporated, allowing efficient exploration of low-probability events that could significantly affect outcomes. The results provide the reliability levels of terminal operation performance metrics under different air traffic intensities, and the sensitivity of reliability to the communication availabilities. It emphasizes implications and guidance for future vectoring procedures in congested terminal areas, highlighting the impact of communication system uncertainties and pilot response delays to the efficiency and safety of civil operations.

4 - Prescriptive Analytics for Optimizing Airport Slot Allocation

Nuno Ribeiro, Singapore University of Technology and Design, Singapore, Singapore

This talk presents recent advancements in prescriptive analytics to support airport slot allocation and the determination of declared capacities. We explore a modeling framework that integrates mixed-integer linear programming (MILP) and heuristic algorithms to optimize slot allocation decisions, while incorporating queuing models to estimate airport delays and machine learning models to predict passenger connections.

SD16

Building A Level 4 A402

Innovations in Fulfillment and Logistics Operations

Invited Session

The Practice Section of Informatics

Chair: SMA Bin Al Islam, Amazon.com, North Reading, MA, 01843, United States

1 - Optimizing Pick Operations Through Distributed Inventory Assignment in Fulfillment Centers

SMA Bin Al Islam, Amazon.com, North Reading, MA, United States, Michael Caldara, Mouhacine Benosman

Amazon's state-of-the-art multilevel containerized inventory system in its next-generation fulfillment center presents new optimization opportunities for customer fulfillment. Where before multi-item customer orders would need to be processed by multiple floors based on inventory availability, our robots now transport containers to and from a vertical lift to process that order on a single floor, thus simplifying downstream fulfillment processes. In this work, we explore several candidate approaches to optimally solve the container relocation problem. We find that traditional centralized optimization models are effective in solution quality; however, when planning across multiple floors, this formulation faces scalability limitations due to time complexity. To address these challenges, we develop a distributed optimization framework that decomposes the global inventory assignment problem into smaller, parallel sub-problems. Each floor-level sub-problem is formulated as a mixed-integer linear program and solved independently, while global coordination is achieved through a Lagrangian relaxation approach. We demonstrate through experimental evaluation that the distributed method offers substantial reductions in runtime while preserving key performance metrics.

2 - Scheduling Jobs for Print on Demand

Alfredo Nantes, Amazon, San Diego, CA, United States, Prem Viswanathan, Yaoguang Zhai

This paper presents a generalized assignment problem (GAP) inspired formulation for unrelated parallel machine scheduling with sequence-dependent setup times. We model the problem using configuration groups that define machine capabilities, reducing decision variables compared to conventional models. Assuming each machine implements configurations from a single group throughout the planning horizon eliminates the need to explicitly model sequence-dependent setups and track individual job completion times, simplifying the model. We apply this formulation to an industrial print optimization problem, demonstrating its effectiveness in balancing customer commitments, costs, and throughput. Experimental results show computational advantages, with the model scaling well and finding near-optimal solutions (relative gap < 1.26%) for large-scale problems of up to 100,000 jobs, 34 machines, and 15 configurations.

3 - Optimizing Delivery Speed Through Shipping Incentives

Prasanna Venkatesh Ramadas, Amazon, Bellevue, WA, United States, Bhaskar Vooradi, Krishna Pendyala

Online shopping experience heavily relies on delivery speed and on-time delivery (OTD), where Fulfillment by Merchant (FBM) sellers handle their own logistics unlike Fulfillment by Amazon (FBA) and often prioritize cost-efficiency over speed. Since faster and reliable deliveries significantly improve customer satisfaction and drive long-term benefits for both Amazon and sellers through increased sales and visibility, shipping discounts are provided as incentives for adopting speedier delivery options. To spend the shipping discount budget optimally for a better customer experience, we build a two-step discount model. The model firstly optimizes the overall delivery speed for all orders by maximizing the expected volume shipped through faster ship methods for a given budget. Then a machine learning model is applied to predict seller affinity on a ship method that is used to continuously refine the estimated expected volume used as input in the optimization model. This approach allows for a more strategic and effective use of shipping discounts, balancing cost considerations with the imperative of rapid delivery in the competitive e-commerce landscape.

SD17

Building A Level 4 A403

Advances in AI for Manufacturing

Panel Session

Practice Track

Co-Chair: Avi Guha, Southwire, TBD, GA, United States

SD18

Building A Level 4 A404

Advances in Traffic Modeling and State Estimation

Invited Session

TSL: Intelligent Transportation Systems

Chair: Ruolin Li, University of Southern California, Los Angeles, CA, United States

1 - Integrating Differential Games of Truck-Borne Freight and Passenger Flows

Terry Friesz, Pennsylvania State University, University Park, PA, United States, Cheng-Chang Lin

The problem we study is the combined congestion of freight (truck) flows and passenger (automobile) flows occurring on the same road network when a portion of passenger vehicles are autonomous. Notably we model freight traffic as a dynamic spatial price equilibrium (DSPE) and passenger traffic as a dynamic user equilibrium (DUE) with superior routing information available to autonomous passenger vehicles (AVs). These modelling perspectives allow the fundamental observation that freight transportation is induced by the spatial separation of production and consumption activities to be directly invoked, while allowing passenger departure time and route choice to occur according to widely endorsed behavioral notions. They also provide the foundation for a model-based digital twin (DT) of urban traffic useful as a tool for congestion forecasting. The challenges of integrating the DSPE and DUE paradigms and setting the stage for a congestion DT will be reviewed, including: (1) finding a recognized mathematical formulation; (2) developing a 2-species (truck and passenger vehicle) path delay operator (PDO); (3) selecting and verifying that one or more algorithms are appropriate to the chosen formulation; (4) demonstrating existence of solutions; and (5) assessing whether the tasks of data assimilation, sensitivity analysis, and near real-time approximate solution can be carried out to an extent that make a digital twin (DT) of combined DSPE/DUE viable.

2 - When Altruism Meets Autonomy: Managing Weaving Ramp Congestion with Strategic AVs

Ruolin Li, University of Southern California, Los Angeles, CA, United States, Kexin Wang, Haohui He

Weaving ramps are critical bottlenecks in highway networks due to conflicting traffic flows and complex interactions among heterogeneous vehicle types. This study seeks to capture, predict, and optimize the aggregate lane-choice behavior of mainline vehicles approaching these zones. We first introduce a lane-choice model for human-driven vehicles (HDVs) based on the Wardrop equilibrium principle, validated through microscopic traffic simulations that demonstrate high predictive accuracy. Building on this foundation, we propose a bi-level Stackelberg-Wardrop framework for mixed traffic, in which connected and autonomous vehicles (CAVs) serve as strategic leaders. At the upper level, CAVs are centrally coordinated to optimize their lane-changing strategies for societal traffic conditions; at the lower level, HDVs adapt via equilibrium-based lane choice. We examine the impact of CAV integration across varying penetration rates and evaluate its effect on overall weaving zone performance. Finally, we explore incentive mechanisms for CAV cooperation by incorporating heterogeneous Social Value Orientations (SVO), highlighting the potential of altruistic behavior and strategic design to enhance social traffic outcomes.

3 - Large-Scale Inverse Learning of User Equilibrium via Multiconvex Optimization

Zhichen Liu, University of Michigan, Ann Arbor, MI, United States, Xi Lin, Yafeng Yin

This study develops a scalable multiconvex optimization framework for inverse learning, from multi-day link flow observations, of contextual travel demand and cost functions for constructing transportation network equilibrium models. The framework parameterizes unknown demand and cost functions using graph-diffusion-based convex bases and formulates the inverse learning problem as a multiconvex optimization. This multiconvex optimization is solved iteratively by addressing two constrained convex subproblems: a quadratic optimization with sparse constraints and a decomposed convex subproblem involving multiple traffic assignment problems. Under the assumption that observed flows are at equilibrium, the framework ensures a tight approximation of the underlying network equilibrium models. By leveraging open-source solvers, the proposed framework promises scalability for modeling large-scale transportation networks using partial link flow observations.

4 - Building Traffic Network Equilibrium Models with Vehicle Trajectory Data

Benjamin Rosenblad, University of Michigan, Ann Arbor, MI, United States, Xi Lin, Yafeng Yin

Transportation network equilibrium models are fundamental tools used in the planning and operation of transportation systems. They help planners evaluate the effectiveness of various improvement strategies and guide decisions on infrastructure investments and policy interventions. Although significant progress has been made in formulating and solving these models, empirical calibration is rather limited, due to the lack of traffic data and computational challenges. However, with the growing availability of vehicle trajectory data made possible by advances in connected vehicle technologies, new opportunities are emerging to overcome these barriers. This research introduces a computationally efficient, end-to-end learning framework that builds traffic equilibrium models directly from vehicle trajectory data. By relying solely on this data source, the approach reduces dependence on traditional inputs from local agencies, making it more adaptable and scalable across cities with varying data resources. Rather than replacing the comprehensive trip-based or activity-based models used in long-range planning, our goal is to develop lightweight models that can support short-term policy decisions and congestion mitigation efforts. By accepting minor trade-offs in behavioral realism for gains in computational speed and scalability, we aim to show that large-scale trajectory data can produce models with sufficient accuracy to meaningfully support decision-making.

SD19

Building A Level 4 A405

Data Driven Control and Optimization of Stochastic Systems

Invited Session

Applied Probability Society

Chair: Andrew Lim, National University of Singapore, Singapore, Singapore

1 - Optimal Sequential Hypothesis Testing with Costly Information Acquisition

Luhao Zhang, Johns Hopkins University, Baltimore, MD, United States

We study the sequential testing problem of two alternative hypotheses when observations are costly. Mathematically, we frame this as an optimal stopping problem in which the decision-maker aims to minimize classification errors plus a running cost proportional to the duration of information acquisition. By formulating a value function that captures both the expected decision risk and the running cost, we derive a set of variational inequalities that govern the evolution of the optimal strategy. Our analysis encompasses two different informational regimes: Gaussian and Poisson. We characterize the continuation regions within which further information collection is optimal and derive monotonicity properties with respect to different parameters, including the noise level, running cost, difference between the two hypotheses, and the relative importance of false positive and false negative errors. For Poisson signals, we utilize a diffusion approximation to study the asymptotic regime of frequent small jumps. Overall, our model advances understanding of how observation costs reshape classical sequential hypothesis testing. It also offers practical guidelines for designing adaptive and non-adaptive testing schemes in medicine, engineering, and related fields under real-world resource constraints.

2 - Optimal Trade Execution with Learning: the Multi-Asset Case

Kok Tong Lim, National University of Singapore, Singapore, Singapore

We consider the problem of optimal trade execution with multiple assets over a finite horizon with a linear permanent price-impact function with parameters that are constant but unknown to the decision maker. The Bayesian stochastic control formulation of this problem is computationally and analytically intractable due to the nonlinearities of the posterior update equation and the dimensionality of the state, which consists of the trader's inventory level and the mean and covariance matrix of all the price-impact coefficients. Using the Hamilton-Jacobi-Bellman dynamic programming equation, we derive an asymptotic expansion of the value function and optimal trading rate in the high-noise regime and characterize the terms of this expansion as solutions of partial differential equations. The zeroth order term corresponds to the "plug-in" (exploitation) decision while higher order terms characterize the exploration strategy and the value of learning. All terms of the expansion can be easily computed by solving a system of recursive equations and play an integral role in the exploration-exploitation trade-off. The higher order terms provide insight into the impact of horizon and inventory on optimal exploration, and an analytical characterization of how to optimally learn interference effects.

3 - On the Bayesian Modeling and Regret Analysis of Nonstationary Bandits

Seungki Min, SNU Business School, Yuseong-Gu, Korea, Republic of

Analyses of nonstationary bandits are often split between two paradigms: adversarial nonstationary models (e.g., unknown sequences with a bounded number of changes or total variation), and stochastic nonstationary models (e.g., autoregressive processes). Adopting the latter approach, this talk presents a general Bayesian regret analysis framework from the paper "An Information-Theoretic Analysis of Nonstationary Bandit Learning", applicable to any stochastic nonstationarity model. Our main result reveals that the intrinsic difficulty of a nonstationary bandit problem is characterized by the entropy rate of the optimal action process. This talk aims to illustrate this key insight through a sharp comparison with established results from the adversarial framework.

4 - Data-Driven Merton's Strategies via Policy Randomization

Yanwei Jia, The Chinese University of Hong Kong, Shatin, Hong Kong, Min Dai, Yuchao Dong, Xunyu Zhou

We study Merton's expected utility maximization problem in an incomplete market, characterized by a factor process in addition to the stock price process, where all the model primitives are unknown. The agent under consideration is a price taker who has access only to the stock and factor value processes and the instantaneous volatility. We propose an auxiliary problem in which the agent can invoke policy randomization according to a specific class of Gaussian distributions, and prove that the mean of its optimal Gaussian policy solves the original Merton problem. With randomized policies, we are in the realm of continuous-time reinforcement learning (RL) recently developed in Wang et al. (2020) and Jia and Zhou (2022a, 2022b, 2023), enabling us to solve the auxiliary problem in a data-driven way without having to estimate the model primitives. Specifically, we establish a policy improvement theorem based on which we design both online and offline actor-critic RL algorithms for learning Merton's strategies. A key insight from this study is that RL in general and policy randomization in particular are useful beyond the purpose for exploration -- they can be employed as a technical tool to solve a problem that cannot be otherwise solved by mere deterministic policies. At last, we carry out both simulation and empirical studies in a stochastic volatility environment to demonstrate the decisive outperformance of the devised RL algorithms in comparison to the conventional model-based, plug-in method.

5 - The Exploration, Exploitation Robustness Tradeoff

Andrew Lim, National University of Singapore, Singapore, Singapore

Parameter uncertainty and model misspecification are fundamental concerns in multi-period data-driven decision making, but little is understood about the tradeoff between exploration, exploitation and robustness. To study this tradeoff, we consider a robust optimal stopping problem with Bayesian learning. We show that an ambiguity-averse decision maker spends less time exploring, even though model uncertainty is reduced by having more data, because the "learning shock" for each posterior update increases the sensitivity of the expected profit to model misspecification. We also show that a willingness to explore can be restored by introducing a hedging instrument that offsets the learning shock.

SD20

Building A Level 4 A406

Advances in Air Transportation

Invited Session

TSL: Air Transportation

Chair: Lavanya Marla, University of Illinois at Urbana-Champaign, Urbana, IL, United States

Co-Chair: Alexander Estes, University of Maryland, College Park, MD, United States

1 - Tri-Level Defender-Attacker-Defender Problem of Bidirectional Energy Supply Logistics

Hyunhwa Kim, University of Illinois Urbana-Champaign, Urbana, IL, United States, Eleftheria Kontou

A cyberattack on the power system could cause severe electricity outages. Sophisticated adversaries attack by magnifying the power load shed by disruption. At the same time, defenders like utility and city planners can harden potentially grid infrastructure in advance to alleviate the damage. During an outage, electric uncrewed aerial (UAVs) and ground (UGVs) vehicles can provide backup power using bidirectional energy exchange technology. These vehicles supply energy to external demand by discharging their batteries as mobile energy sources and recharge at charging infrastructures if needed. However, the functionality of charging infrastructure depends on the power system disruption because the required charging energy is transmitted through the power grid. We develop a tri-level defender-attacker-defender problem to provide a planning strategy for outages caused by cyberattacks. This problem integrates transportation and power systems with charging infrastructure connecting both networks. The problem consists of: (1) a fortification framework to minimize power load shed and travel distance by fortifying power lines; (2) an interdiction framework to maximize power load shed by damaging lines under limited attacker budget; and (3) an operational framework to determine UAV and UGV routes and charging station locations. The problem is formulated as a min-max-min model and decomposed into two bi-level problems. The defender's problem is solved using logic-based Benders' decomposition and adaptive large neighborhood search. The attacker's problem is reformulated into a single-level binary linear program. This model is applied to integrated transportation and power networks, providing planning insights as a response to coupled transport-power system cyberattacks.

2 - Decentralized Coordination of Autonomous Traffic Through Advanced Air Mobility Corridors

Jasmine Jerry Aloor, Massachusetts Institute of Technology, Cambridge, MA, United States, Hamsa Balakrishnan

"The use of dedicated corridors for Advanced Air Mobility (AAM) traffic is one of the most commonly proposed pathways to integrating them into existing airspace operations. Most prior research has focused on the design of networks of AAM corridors and conflict resolution for aircraft within corridors. It is also generally believed that while attractive from an implementation perspective, corridor-based operations may be inefficient, especially in the absence of centralized traffic management. In this paper, we show that contrary to this belief, it is possible for autonomous aircraft to learn to self-organize into corridor flows in decentralized settings. We illustrate our approach using scenarios in which fixed-wing aircraft need to safely and efficiently traverse (1) a single corridor with metering after the exit, (2) a sequence of two consecutive corridors, and (3) a corridor that splits into two. We find that in decentralized settings with only local information, the aircraft are able to conform to the corridor boundaries more than 95% of the time and reach their goal in a relatively efficient manner. Furthermore, tactical interventions to handle violations of the separation minimum are needed only infrequently in low- and medium-density settings. However, such tactical interventions become more frequently necessary only when traffic density is high."

3 - A Dynamic Model for Airline Fleeting and Scheduling

Jingxing Wang, University of California Berkeley, Berkeley, CA, United States, Chiwei Yan, Archis Ghate

COVID-19 has reshaped the global airline industry. Travel demands are volatile, and passengers have more flexibility in bookings and cancellations. More than ever, airlines have to be agile and adaptive while making operational decisions. We introduce a dynamic model to help airlines make adaptive fleeting and scheduling decisions based on stochastically evolving bookings. The model incorporates both arrivals (new bookings) and departures (cancellations). This significantly generalizes and strengthens previous modeling attempts. We develop a Lagrangian relaxation framework that decomposes the dynamic program defined on a large time-space network into separable flight-level problems. Owing to the complexities of this network, our Lagrangian dual problem is not straightforward to solve. We therefore develop a tailored yet simple projected subgradient algorithm that exploits the structure of this network for efficient solutions. We establish the correctness and convergence properties of this procedure. Our analysis yields new theoretical, algorithmic, and managerial insights into the dynamic fleeting and scheduling problem. We present computational experiments based on real-world airline data to demonstrate the potential benefits of this approach.

4 - Aircraft Trajectory Decisions for Combinatorial Sampling in the National Airspace

Lavanya Marla, University of Illinois at Urbana-Champaign, Urbana, IL, United States

We study the value of collecting information in highly non-stationary settings with spatial correlations and information decay, motivated by the lack of near real-time information for routing aircraft. We propose to use a fleet of en-route aircraft as weather "sensors" to sample and collect near-real-time information so as to reduce long-term uncertainty about the airspace. We develop a stylized graph structure to capture the combinatorial nature of realistic aircraft trajectory networks, and model the cost incurred due to weather and wind over each arc using a unique spatial-temporal correlation structure. We mathematically show how to optimally explore and collect information to maximize information gain first in temporally correlated environments and next in jointly spatially and temporally correlated environments under combinatorial network structures. We demonstrate first that it is possible to collect information by parsimonious but careful sampling of the airspace. Next we show that the fuel burn reduction due to this additional information is significant, resulting in reduced environmental impact of aviation.

5 - Strategic Capacity Planning for a Network of Aircraft Maintenance Facilities

JohnPaul Adimonyemma, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Yanshuo Sun

Most existing studies in aircraft maintenance optimization have centered on check scheduling, often overlooking the critical role of capacity planning for maintenance stations. To minimize total maintenance costs, airlines must jointly optimize facility capacity and maintenance scheduling based on projected aircraft utilization. This study addresses that gap by proposing a novel integrated mixed-integer quadratic program optimization model that jointly optimizes station capacities, check scheduling, and aircraft-to-station assignments. This model accounts for fleet heterogeneity, station-specific capabilities, stations' capacity limits, and projected aircraft utilization with the aim of minimizing total maintenance cost. Due to the complexity of large instances, traditional solvers are impractical. Hence, we propose decomposition algorithms, including one augmented with machine learning. A major U.S. airline case study will be presented, which will enable airlines and professionals to properly plan for their maintenance operations.

6 - Rethinking Crew Pairing Optimization: A GPU Accelerated Approach

Filipe Cabral, American Airlines, Inc, Fort Worth, TX, United States, Yanqing Kuang

Airlines heavily rely on the Crew Pairing Problem (CPP) to minimize crew costs by creating flight sequences (pairings) that cover all scheduled flights while adhering to complex operational and regulatory constraints. Formulated as a set partitioning problem, CPP's vast solution space challenges traditional methods like branch-and-price and column generation to their computational limits. These algorithms can even become computationally infeasible for large legacy airlines. Our research approaches this challenge by leveraging the power of GPUs, the very hardware behind the recent success of Large Language Models and Deep Neural Networks. Specifically, we shift from the solution of computationally expensive linear systems inherent in Simplex and Interior Point methods to a more scalable approach using a first-order method called cuPDLP. Preliminary results using real-world airline data demonstrate significant performance gains over traditional CPU-based approaches, highlighting the potential of our method.

SD21

Building A Level 4 A407

Learning and Experimentation

Invited Session

Revenue Management

Chair: Will Ma, Columbia Business School, New York, United States

Co-Chair: Jinglong Zhao, Boston University, Boston, MA, United States

1 - Experiment Design for Assortment Optimization

Xintong Yu, Columbia Business School DRO, NYC, NY, United States, Will Ma

We propose a structured experimental design framework for assortment optimization, aiming to improve parameter estimation across a wide range of consumer choice models. While discrete choice models are well-studied, how to *design experiments*—that is, which assortments to offer to best learn preferences—remains underexplored.

Our work makes two main contributions:

Deterministic Learning of Nest Structures: In nested logit models, products are grouped into latent “nests” reflecting correlated preferences. Typically assumed known, these nests are often unknown in practice. We show that, under exact choice probabilities, a simple and scalable non-adaptive experimental plan can recover the true nest structure using only a small number of queries. This solves a long-standing open problem.

Efficient Choice Estimation: Beyond nested logit, we evaluate our design on general models like MNL, Markov, and ranked list models. Compared to random assortment designs, our fixed plan achieves lower estimation error under equal sample budgets. Built on the Random Utility Model framework and leveraging existing estimation tools, our results isolate the benefit of design alone.

Overall, our framework offers both theoretical guarantees and strong empirical performance. It is especially impactful in large-scale settings—like e-commerce or fantasy sports—where choosing the right product mix is crucial. Our results demonstrate that strategic experimentation, rather than random trial-and-error, can significantly enhance learning of consumer behavior.

2 - A Generalized Bayesian Framework for Online Optimization

Kaizheng Wang, Columbia University, New York, NY, United States

We introduce a generalized Bayesian framework for online optimization, centered on a new concept of generalized posterior distribution that effectively summarizes uncertainty from partial knowledge. This approach extends existing Bayesian methods while offering enhanced flexibility to incorporate valuable prior knowledge, such as shape constraints on the objective function. New bandit problems derived from this framework are shown to achieve optimal regret bounds up to logarithmic factors.

3 - Data-Driven Sequential Search

Cagin Uru, Duke University, Durham, NC, United States, David Brown

How can one search effectively with limited prior knowledge on the distribution of alternative values? This paper addresses this question in the context of sequential search. In our data-driven setting, traditional methods that rely on a known value distribution no longer apply and feasible search policies are based solely on the history of explored alternative values. We seek to identify a policy that maximizes the worst-case ratio of expected reward compared to an oracle (referred to as Pandora) with full knowledge of the value distribution. We first consider a general setting where alternative values may follow any arbitrary distribution. We design simple policies that statically commit to a prespecified number of explorations. We show that these static policies guarantee a competitive ratio of at least $1/e \approx 37\%$ of the Pandora benchmark for any arbitrary value distribution. Our approach involves studying nature's problem to select a distribution to counter a policy and identifying worst-case distributions. Moreover, we study how the structure of the unknown value distribution influences achievable performance guarantees by considering a setting where feasible distributions belong to the class of monotone hazard rate distributions. In this case, we improve our guarantee to $(e/(e+1))^2 \approx 53\%$ of the Pandora benchmark. In both settings, we derive performance limits for feasible policies to complement our performance guarantee results. We illustrate the efficacy of our policies through numerical examples based on price search. Our results provide insights to the decision makers with no prior knowledge of value distributions across many search applications.

4 - Gen-DFL: Decision-Focused Generative Learning for Robust Decision Making

Prince Wang, Carnegie Mellon University, Pittsburgh, PA, United States

This talk introduces Gen-DFL, a framework that enhances traditional decision-focused learning by integrating generative models to adaptively sample uncertainty. Gen-DFL improves robustness in high-dimensional and risk-sensitive environments while maintaining interpretability and performance. Applications in logistics and scheduling highlight strong gains over baseline methods.

SD22

Building A Level 4 A408

Self-Exciting Processes

Invited Session

Applied Probability Society

Chair: Andrew Daw, University of Southern California, Los Angeles, United States

1 - Single-Server Queues with State-Dependent Hawkes Arrivals

Guodong Pang, Rice University, Houston, TX, United States, Bo Li

We study single-server queues with Hawkes arrivals whose intensity process depends on the queue length or workload through the self-exciting function, and with i.i.d. general service times, under the first-come first-served discipline. We prove the functional law of large numbers (FLLNs) and functional central limit theorems (FCLTs) for the joint processes of the arrivals, queue-length and workload processes, in the heavy traffic regime. The fluid limit is given by a set of nonlinear integral equations such that the fluid queue or workload has a reflection at zero. We analyze the transient and equilibrium behaviors of the fluid limit, in particular, identifying the equilibrium points for the queue or workload fluid model. We assume that the fluid limit is at an equilibrium point in order to establish the FCLTs for the joint diffusion-scaled arrival, queue-length and workload processes in the critically load regime. When the equilibrium point of the queue is at zero, the limit for the joint processes satisfies a stochastic differential equation such that the queue-length or workload limit has a reflection at zero. In particular, the queue-length or workload process is equivalent in distribution to a generalized (possibly nonlinear drift) Ornstein-Uhlenbeck (OU) diffusion with reflection at zero. When the equilibrium point is positive, the limit for the joint processes is a diffusion process without reflection, in particular, the queue-length or workload process is a generalized OU diffusion.

2 - Spatiotemporal Hawkes Processes with a Graphon-Induced Connectivity Structure

Justin Baars, Universiteit van Amsterdam, Amsterdam, Netherlands, Roger Laeven, Michel Mandjes

We introduce a spatiotemporal self-exciting point process $\{N_t(x)\}$, boundedly finite both over time $[0, \infty)$ and space $\mathbb{X}$, with excitation structure determined by a graphon W on $\mathbb{X}^2$.

This graphon Hawkes process generalizes both the multivariate Hawkes process and the Hawkes process on a countable network, and despite being infinite-dimensional, it is surprisingly tractable.

After proving existence, uniqueness and stability results, we show, both in the annealed and in the quenched case, that for compact, Euclidean $\mathbb{X} \subset \mathbb{R}^m$, any graphon Hawkes process can be obtained as the suitable limit of d -dimensional Hawkes processes $\tilde{N}^d$, as $d \rightarrow \infty$.

Furthermore, in the stable regime, we establish an FLLN and an FCLT for our infinite-dimensional process on compact $\mathbb{X} \subset \mathbb{R}^m$, while in the unstable regime we prove divergence of $N_T(\mathbb{X})/T$, as $T \rightarrow \infty$.

Finally, we exploit a cluster representation to derive fixed-point equations for the Laplace functional of N , for which we set up a recursive approximation procedure. We apply these results to show that, starting with multivariate Hawkes processes $\tilde{N}^d_t$ converging to stable graphon Hawkes processes, the limits $d \rightarrow \infty$ and $t \rightarrow \infty$ commute.

3 - Multivariate Generalized Polya Processes and Applications

Brian Fralix, Clemson University, Clemson, SC, United States, Jeffrey Kharoufeh, David Pittman

We will examine various types of multivariate Generalized Polya Processes, as well as various stochastic models “fed” by these type of point processes. Connections to other models found within the literature on Polya urn processes will also be discussed.

4 - Dynamic Pricing Under Self-Exciting Arrival Processes

Longyuan Du, University of San Francisco, San Francisco, CA, United States, Quan Yuan, Ming Hu

We study dynamic pricing under a stochastic, non-linear, self-exciting demand arrival process over a finite sales horizon. We adopt such a correlated demand process to capture the phenomenon that customers who have made a purchase can inform and excite future customers to arrive. Specifically, the stochastic arrival intensity is

stimulated with an immediate boost after any purchase and gradually decays over time between two consecutive purchases. The arrival intensity is also affected by the market phase the seller operates in, a growth or maturity stage, which convexly or concavely shapes how many new customers informed customers can stimulate. We show that the optimal pricing policy depends on the current time and excitement level, which measures the currently accumulated influence that previously informed customers exert and is shown to be the only sufficient statistic for our problem. Over time, the seller tends to initially set a low price to attract early customers who can excite later arrivals and later increase the price to maximize profits. Perhaps surprisingly, for a given time, in the growth stage of the market, the optimal price rises with the excitement level, and conversely, in the maturity stage of the market, the optimal price drops with the excitement level. Moreover, we show that adopting the optimal time-varying (open-loop) pricing policy solved from the deterministic problem leads to only a constant performance loss from the optimal dynamic (i.e., contingent) pricing policy, which would not hold for the classic revenue management problem with a capacity constraint.

5 - Stopping Self-Excitement

Andrew Daw, University of Southern California, Los Angeles, CA, United States

From management of service interactions to studies of neural signal propagation, there is a subtle but fundamental challenge that comes with such endogenously spurred activity: when should you say it's over? If each point has the potential to beget future points, there is an inherent tradeoff between omission and excess: stop too early and you may miss further observations, but stop too late and you've wasted too much time. In this paper, we study an optimal stopping problem that aims to balance this omission-excess tradeoff within a cluster of history-driven events. We analyze and prescribe stopping policies for the self-exciting stochastic process through a transformation to the workload process for an M/D/1 queueing system. Then, to connect these model-based insights to practice, we draw upon ideas from the algorithms-with-predictions literature and improve the model-based policies to adapt to contextual information, such as the texts within a service interaction. Finally, we demonstrate the performance of these policies through application to a publicly available linguistics data set.

SD23

Building A Level 4 A409

Driving Equity in Data-Informed Decisions

Panel Session

Minority Issues Forum

Co-Chair: Xiaoyan (Amber) Liu, Santa Clara University, Santa Clara, CA, United States

Co-Chair: Hussein El Hajj, Santa Clara University, Santa Clara, CA, 95053, United States

1 - Moderator Panelist

Hussein El Hajj, Santa Clara University, Santa Clara, CA, United States

As data-driven decision-making becomes increasingly central to operations across sectors, it is imperative to examine who benefits from these decisions, and who may be left behind. This panel explores how data, algorithms, and analytics can both perpetuate and address systemic inequities. We will discuss frameworks, case studies, and methodologies that prioritize equity in design, implementation, and evaluation.

2 - Panelist

Kejia Hu, University of Oxford Said Business School, Oxford, United Kingdom

3 - Panelist

Violet (Xinying) Chen, Stevens Institute of Technology, Hoboken, NJ, United States

4 - Panelist

Abdullah Konak, Penn State Berks, Reading, PA, United States

5 - Moderator Panelist

Xiaoyan (Amber) Liu, Santa Clara University, Santa Clara, CA, United States

As data-driven decision-making becomes increasingly central to operations across sectors, it is imperative to examine who benefits from these decisions, and who may be left behind. This panel explores how data, algorithms, and analytics can both perpetuate and address systemic inequities. We will discuss frameworks, case studies, and methodologies that prioritize equity in design, implementation, and evaluation.

SD24

Building A Level 4 A410

Solving Quadratic Unconstrained Binary Optimization Models on Quantum Computers"

Tutorial session

Tutorials

Chair: Nilay Noyan, AMAZON, Bellevue, WA, United States

1 - Solving Quadratic Unconstrained Binary Optimization Models on Quantum Computers

Sridhar Tayur, Carnegie Mellon University, Pittsburgh, PA, United States, Arut Rhik Mazumder

This tutorial offers a quick, hands-on introduction to solving Quadratic Unconstrained Binary Optimization (QUBO) models on currently available quantum computers and their simulators. We cover both IBM and D-Wave machines: IBM utilizes a gate-circuit architecture, and D-Wave is quantum annealer. We provide examples of three canonical problems and two models from practical applications. The tutorial is structured to bridge the gap between theory and practice: we begin with an overview of

QUBOs, explain their relevance and connection to quantum algorithms, introduce key quantum computing concepts, provide the foundations for two quantum heuristics, and provide detailed implementation guides. An associated GitHub repository provides the

codes in five companion notebooks. In addition to reaching undergraduate and graduate students in computationally intensive disciplines, this article aims to reach working industry professionals seeking to explore the potential of near-term quantum applications.

As our title indicates, this tutorial is intended to be a starting point in a journey towards solving more complex QUBOs on quantum computers.

SD25

Building A Level 4 A411

Life After the PhD: Exploring Careers in Industry and Academia

Panel Session

Women in OR/MS (WORMS)

Co-Chair: Irem Sengul Orgut, Loyola University Chicago, Tuscaloosa, AL, United States

1 - Moderator Panelist

Irem Sengul Orgut, Loyola University Chicago, Chicago, IL, United States

TBD

2 - Panelist

Lauren Davis, North Carolina A&T State University, Greensboro, NC, United States

3 - Panelist

Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States

4 - Panelist

Cara Touretzky, Gurobi Optimization, Santa Fe, NM, United States

5 - Panelist

Masoud Barah, W.W. Grainger, Lake Forest, IL, United States

SD26

Building A Level 4 A412

Emerging Topics in Location and Logistics

Invited Session

Location Analysis

Chair: Sibel Alumur Alev, University of Waterloo, Waterloo, ON, Canada

1 - Demand-Driven Hub and Service Network Design Under Uncertainty

Hao Li, FICO, Oakville, ON, Canada

We consider a planning problem for freight transportation carriers that seeks to profitably match supply with demand while recognizing uncertainty in shipment volumes. On the supply side, the problem determines transportation network design decisions regarding hub locations and the number of vehicles to be dispatched within the network in each period of the planning horizon. On the demand side, the problem incorporates the carrier's ability to expand its service coverage by selectively accepting additional customer demands beyond its existing contractual base. Furthermore, while some of these additional customers seek a long-term commitment from the carrier, others are transactional and only require the transportation of a single set of shipments. We refer to this problem as the Demand-driven Hub and Service Network Design under Uncertainty problem and formulate it as a two-stage stochastic program. Further, we develop an enhanced Benders Decomposition-based solution method for solving instances of this model. The solution methodology is inspired by Partial Benders Decomposition, leveraging a problem reformulation that embeds subsets of subproblem variables and constraints into the master problem, while also incorporating valid inequalities to strengthen the formulation. We illustrate with an extensive computational study that the proposed method outperforms adaptations of benchmarks proposed for similar problems. We validate the benefits of solving the proposed model, which integrates decisions that have not yet been jointly modeled, with an analysis based on Sample Average Approximation.

2 - Electrification of Freight Transportation Hub Networks with Emissions-Sensitive Customers

Gita Taherkhani, Loyola University Chicago, Chicago, IL, United States, Dipayan Banerjee, Mike hewitt

This paper investigates a tactical and strategic planning problem faced by a motor carrier operating a fixed middle-mile hub network for clients sensitive to greenhouse gas (GHG) emissions. Commodities purchased by retailers from multiple vendors must be transported from their origins to destinations via a hub-to-hub network, with all-or-nothing service decisions. The carrier must decide which commodities to serve, how to route accepted shipments, where to install electric vehicle (EV) charging infrastructure, and how many diesel and electric trucks to deploy on each lane, all while satisfying client-specific GHG emission budgets. Emissions are estimated using the GLEC well-to-wheel framework and allocated proportionally among shippers. To support these decisions, we develop a mixed-integer programming formulation for hub network design. Using realistic instances, we conduct an extensive computational study on both deterministic and stochastic versions of the problem. The analysis aims to (1) evaluate the computational performance of the proposed model and (2) examine the impact of incorporating GHG emission constraints on optimal network design.

3 - Pattern-Based Kernel Search for the Single-Source Capacitated Facility Location Problem

Stefan Nickel, Karlsruhe Institute of Technology, Karlsruhe, Germany, Hannah Bakker, Gianfranco Guastaroba, Maria Grazia Speranza

We introduce Pattern-based Kernel Search (PaKS), a two-phase metaheuristic expanding the successful Kernel Search (KS) for the capacitated facility location problem with single sourcing (SSCFLP). KS leverages the fact that solving a sequence of small BIPs can be more efficient than solving one large BIP by partitioning decision variables into disjoint subsets — a kernel of “promising variables” and complementary buckets. An initial solution is obtained by solving the restricted BIP with only the kernel variables. Then, a sequence of restricted BIPs combining kernel and bucket variables is solved, progressively improving the solution and updating the kernel. The performance of kernel search is critically impacted by the partitioning of the variables into the kernel and the buckets. While KS relies solely on information from the LP relaxation, PaKS uses pattern recognition to analyze a series of modified LP relaxations in its first phase. As a result, facilities and customers are divided into regions, each region corresponding to a bucket. Leveraging this implied spatial information allows to drastically reduce the size of the buckets. For the SS-CFLP, the largest instances solved in the literature contain up to 1,000 facilities and customers each. Computational experiments show that PaKS outperforms state-of-the-art approaches even on benchmark instances containing up to 2,000 facilities and customers each, demonstrating its strong potential for solving large-scale problems.

4 - Handling Volunteer Convergence from An OR Perspective: Volunteer Based Team Formation

Bahar Yetis Kara, Ihsan Dogramaci Bilkent University, Ankara, Turkey, Egehan Ugras, Zehranaz Donmez

Search and Rescue (SAR) operations are essential in disaster response, particularly after earthquakes where immediate action saves lives. Their success depends on effective team management and coordination. Urban Search and Rescue (USAR) teams, specialized in rescuing people from collapsed structures, are classified by the International Search and Rescue Advisory Group (INSARAG) into three levels: Light, Medium, and Heavy. These multi-disciplinary teams are crucial for complex disaster environments. Our study presents a mathematical model for the formation and assignment of SAR teams. The model assigns volunteers to teams, forms teams with appropriate expertise, and allocates them to debris sites while considering operational constraints. The model incorporates debris site requirements and the feasibility of assigning specific team types to different debris locations. A key feature of our approach is the introduction of an additional team category between light and medium teams, designed to improve flexibility in SAR operations. This new team type allows for partially completed teams that can either be combined with international teams for full functionality or provide critical information on missing personnel needed to meet medium team requirements. The model identifies gaps in staffing and allows this information to be shared with professional chambers, NGOs, or relevant organizations to recruit the necessary personnel or adjust assignments as new volunteers arrive. Model is also applicable for different disaster scenarios. By incorporating this adaptive team structure, our model enhances coordination in disaster response, ensuring that SAR efforts can be dynamically adjusted based on resource availability.

SD27

Building B Level 2 B201

Theory and Applications in Stochastic Integer Programming

Invited Session

Computing Society

Chair: Haoyun Deng, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Weijun Xie, Georgia Institute of Technology, Atlanta, GA, United States

1 - Mobile Parcel Locker Scheduling with Customer Choices Under Uncertain Demand

Can Yin, University of Minnesota, Twin Cities, Minneapolis, MN, United States, Yiling Zhang

The rapid growth of e-commerce has increased demand for fast delivery, straining last-mile logistics. Additionally, traditional home deliveries often lead to missed packages and theft, raising costs and inefficiencies, while fixed parcel locker locations limit accessibility, especially in low-density areas. To overcome this, mobile parcel lockers provide a flexible alternative by dynamically relocating based on demand to improve efficiency and reduce customer travel. However, their implementation faces challenges, including responding to uncertain customer demand preferences, and balancing the interplay between long-term planning. Existing research largely emphasizes the latter challenge but overlooks the impact of dynamic and uncertain demand patterns. Our work addresses these gaps by proposing an integrated framework to guide the planning and scheduling of mobile parcel locker systems. In particular, first, we account for the interdependencies between customer demand and locker schedule and locations using a variant of the multinomial logit (MNL) model to represent customer preferences among different available options. Second, we consider the interactions between long-term scheduling decisions and operational capacity allocation under demand uncertainty following unknown distributions. The problem is formulated as a two-stage mixed 0-1 distributionally robust optimization (DRO) model. To solve the proposed models efficiently, we exploit the hidden discrete dual structure of the problem, identify the worst case distribution based on the supermodularity of second stage problem. Later, we propose an equivalent mixed-integer linear programming (MILP) reformulation. Additionally, we develop practical variants for flat-rate and locational pricing policies to accommodate real-world deployment.

2 - Stochastic Optimization Models for the Generalized Assignment Problem

Harshit Kothari, University of Wisconsin-Madison, Madison, WI, United States, Jim Luedtke, Luke Marshall, Konstantina Mellou

The Generalized Assignment Problem (GAP) involves assigning tasks to agents with capacity constraints, aiming to minimize total cost while respecting resource limits. In many real-world settings, uncertainty in task requirements and agent capacities introduces stochasticity, giving rise to the Stochastic Generalized Assignment Problem (SGAP). This talk addresses a variant of SGAP relevant to cloud service providers, where virtual machines (jobs) are assigned to physical machines under uncertainty in job demands and arrivals. We formulate this as a two-stage stochastic integer program: in the first stage, initial jobs are assigned, and in the second stage, additional jobs arrive with uncertain resource requirements and costs. First-stage assignments may be relocated, incurring relocation penalties. Due to the complexity of solving SGAP with integer recourse, we explore decomposition approaches like column generation.

3 - Dynamic Inspection Routing in Infrastructure Networks with Decision-Dependent Cascading Failures

Juan-Alberto Estrada-Garcia, University of Michigan, Ann Arbor, MI, United States, Siqian Shen

We study the problem of dynamically scheduling inspection routes in infrastructure systems, such as water distribution and power networks, under uncertainty and interdependent risks. These systems are characterized by aging components whose failures can propagate, triggering cascading disruptions across the system. We formulate a multistage stochastic dynamic integer program, where component failure probabilities evolve in a decision-dependent manner, and cascading effects are explicitly represented. To address the computational challenge, we develop a solution approach based on the Stochastic Dual Dynamic Integer Programming (SDDIP) algorithm and benchmark its performance against a Deep Q-learning method from the literature. We focus our evaluation on water distribution networks, whose mesh-like structure and operational interdependencies provide a realistic setting for analyzing the impact of routing decisions on cascading failure risk.

4 - Scalable and Robust in-Database Stochastic Optimization

Anh Mai, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

In-database stochastic optimization often involves choosing a multi-set of tuples (row in a relational database table) that optimizes an expected outcome while minimizing risks. Sample Average Approximation (SAA) uses millions of scenarios (realizations of stochastic attributes of all the tuples) to formulate an Integer Linear Program (ILP). This ILP contains millions of decision variables corresponding to the number of tuples in the database and a large set of artificial variables and constraints corresponding to the number of scenarios needed to reasonably approximate the optimal solution. On the other hand, SummarySearch formulates its ILP after reducing the number of constraints by summarizing the large set of scenarios into a much smaller set of conservative summaries. In this way, SummarySearch aims to achieve its scalability by searching for the fewest summaries that return a near-optimal solution. Nevertheless, the solutions found by SummarySearch or SAA lack robustness when the number of scenarios is limited and the empirical distribution is inaccurate. In contrast, Robust Satisficing (RS) directly optimizes a solution that reliably attains a target objective even when the true data distribution deviates significantly from the empirical one. However, RS lacks scalability due to having the same ILP size as SAA. In this work, we explore a hybrid algorithm that adapts RS for solving stochastic optimization in database while making it scalable via the scenario-summarization technique introduced in SummarySearch and via a novel variable pruning strategy.

5 - Stochastic Bilevel Optimization for Network Design of Multimodal Transit Systems

Suri Liu, Dalian University of Technology, Dalian, China, People's Republic of, Yiling Zhang, Beste Basciftci, Wenyan Wang

This paper investigates the network design problem of the multimodal transit system under travel time uncertainty. A stochastic bilevel model is proposed to capture the strategic interaction between the transit agency and the riders. The objective is to design a bus route network that balances the interests of both agents and achieves an equilibrium. A tree search algorithm is proposed to solve the large-sized bilevel model, which enumerates the follower responses and represents their behavior using a set of linear constraints, thereby converting the bilevel model to a single-level formulation. To address the computational intractability of exhaustively enumerating the responses of all followers, a Column-and-Constraint Generation algorithm is developed to obtain the optimal solution efficiently. The superiority of the proposed algorithm in comparison with existing single-level reformulations is evaluated on a small-sized synthetic dataset. Finally, the approach is applied to a real-world network design case, providing an analysis of the benefits of adopting the multimodal transit system.

SD28

Building B Level 2 B202

Optimization of Power Systems

Invited Session

Computing Society

Chair: Daniel Bienstock, Columbia University, New York, NY, 10027, United States

1 - Electricity Markets with Consumer-Defined Carbon Costs: a New Paradigm for Greening the Grid

Wenqian Jiang, University of Wisconsin-Madison, Madison, WI, United States

An increasing number of electricity consumers can be characterized as carbon-sensitive, i.e. they are willing to adapt when (and in some cases where) they consume electricity based on the real-time carbon intensity of electricity. However, current electricity markets provide limited opportunity for consumers to express their preferences for low-carbon electricity. To provide consumers, ranging from carbon-agnostic to carbon-sensitive, with a tool to actively manage their roles in carbon emission mitigation, we propose a new green electricity market clearing paradigm, where consumers can not only bid for power but also assign a cost to the carbon emissions incurred by their electricity use. The centralized market clearing problem maximizes social welfare under consideration of carbon preferences, and from this problem, we derive an equivalent equilibrium formulation. Based on the equilibrium formulation, we show that differences in consumers' carbon costs and generators' carbon emission intensities result in carbon-adjusted electricity prices and prove that the proposed market clearing satisfies revenue adequacy and cost recovery. Our investigation informs the ongoing debate about carbon-aware market design in the electricity sector. Joint work with Line Roald.

2 - Improving the Accuracy of the DC Power Flow Approximation via Parameter Optimization

Daniel Molzahn, Georgia Institute of Technology, Atlanta, GA, United States, Babak Taheri

Many power system operation and planning problems use the DC power flow approximation to address computational challenges from the nonlinearity of the AC power flow equations. The DC power flow simplifies the AC power flow equations to a linear form that relates active power flows to phase angle differences across branches,

parameterized by coefficients based on the branches' susceptances. Inspired by techniques for training machine learning models and recent progress in differentiable optimization, this talk presents several algorithms that seek optimal coefficient and bias parameters to improve the DC power flow approximation's accuracy for various applications, including optimal power flow and optimal transmission switching.

3 - Dual Conic Proxy for Semidefinite Relaxation of AC Optimal Power Flow

Qiu Guancheng, Georgia Institute of Technology, Atlanta, GA, United States

The nonlinear, non-convex AC Optimal Power Flow (AC-OPF) problem is fundamental for power systems operations. The intrinsic complexity of AC-OPF has fueled a growing interest in the development of optimization proxies for the problem, i.e., machine learning models that predict high-quality, close-to-optimal solutions. More recently, dual conic proxy architectures have been proposed, which combine machine learning and convex relaxations of AC-OPF, to provide valid certificates of optimality using learning-based methods. Building on this methodology, this paper proposes, for the first time, a dual conic proxy architecture for the semidefinite (SDP) relaxation of AC-OPF problems. Although the SDP relaxation is stronger than the second-order cone relaxation considered in previous work, its practical use has been hindered by its computational cost. The proposed method combines a neural network with a differentiable dual completion strategy that leverages the structure of the dual SDP problem. This approach guarantees dual feasibility, and therefore valid dual bounds, while providing orders of magnitude of speedups compared to interior-point algorithms. The paper also leverages self-supervised learning, which alleviates the need for time-consuming data generation and allows to train the proposed models efficiently. Numerical experiments are presented on several power grid benchmarks with up to 500 buses. The results demonstrate that the proposed SDP-based proxies can outperform weaker conic relaxations, while providing several orders of magnitude speedups compared to a state-of-the-art interior-point SDP solver.

4 - Understanding and Mitigating Real-TIME Price Spikes in Power Markets

Alexander Derenchuk, Colorado School of Mines, Golden, CO, United States, Tomas Tapia, Yury Dvorkin, Daniel Bienstock, Alexandra Newman

In wholesale power markets, system operators use real-time balancing auctions to deploy and compensate additional generation necessary to account for deviations between expected and realized demand (i.e., volatility). Often, prices in these real-time markets greatly exceed the marginal cost of any individual power plant contributing generation to meet load. These price "spikes" can be caused by systemwide generation scarcity, which triggers administratively set prices that activate when there is a shortage of operating reserves. They can also be caused by local generation scarcity, which manifests as transmission network congestion. Under congestion, high prices can arise from penalties for exceeding flow limits on transmission lines in the economic dispatch algorithm used by the system operator. Increasing penetration of solar and wind generation in power systems can increase volatility and thus the likelihood of real-time systemwide scarcity. It can also increase on the system due to the typical placement of large wind and solar power plants in rural areas far from urban centers with high power demand. Using data from Oklahoma Gas and Electric (OGE) and the Southwest Power Pool (SPP), we identify the causes, locations, and timing of price spikes in the SPP grid region and analyze their potential impact on member utilities such as OGE. We then investigate whether strategic battery storage deployment at specific grid locations in critical periods could dampen or eliminate these spikes, and we identify optimal policies for battery storage deployment and operation. Joint work with Tomas Tapia, Yury Dvorkin, Alexandra Newman and Daniel Bienstock.

5 - Linear Relaxations for Large-Scale Acopf

Matias Villagra, Columbia University, New York City, NY, United States, Daniel Bienstock

In this talk, we present fast and scalable algorithms for computing dual bounds of challenging Multi-Period Security Constrained AC-OPF instances from the ARPA-E Grid Optimization Competition. We primarily focus on linear relaxations that are dynamically computed. We present results on very large instances. Joint work with Daniel Bienstock.

SD29

Building B Level 2 B203

Improving Coordination of Interdependent Natural Gas and Power Systems during Winter Storms: Risks and Impacts

Invited Session

ENRE: Energy

Chair: Chiara Lo Prete, The Pennsylvania State University John and Willie Leone Family Department of Energy and Mineral Engi, University Park, PA, United States

1 - Predicting Gas-Fired Generator Performance Under Extreme Winter Weather in the Northeastern United States

Anamika Dubey, Washington State University, Pullman, WA, United States

Extreme winter weather events pose significant risks to the operation of gas-fired power generators. We utilize generator outage data from the NERC GADS, ambient temperature records from NOAA, and system demand data from NYISO and ISO New England in a two-part Bayesian regression model to predict the derated output of gas-fired generators in the Northeastern United States under extreme freezing conditions. The first stage employs Bayesian logistic regression to estimate the probability of a non-zero (derated) output, while the second stage applies Bayesian linear regression to predict the derated capacity conditioned upon non-outage scenarios. Results show that generators exhibit increasing probabilities of complete outages with decreasing temperature and increasing system demand. Although the model is currently limited by the aggregation of temperature data at the state level and incomplete event cause labelling, it provides a probabilistic framework to assess generator net available capacity during extreme winter scenarios. By leveraging this framework, system operators can identify most vulnerable generators to extreme cold weather and develop targeted planning strategies to enhance overall grid resilience. Access to more granular data, including generator-specific locations and accurately labelled outage causes, is necessary for improving model predictability and reducing uncertainty.

2 - Quantifying Winter Storm Risks on Interdependent Power and Natural Gas Systems

Malo Lahogue, Massachusetts Institute of Technology, Cambridge, MA, United States, Saurabh Amin, Rahman Khorramfar

Extreme winter storms increasingly threaten the reliability of interdependent natural gas and electric power systems by simultaneously disrupting gas-fired generation, reducing fuel availability, and increasing heating and electricity demand. As climate volatility intensifies, system operators and planners need tools to quantify the risk of such rare, high-impact events and to guide the design of more resilient systems. Conventional reliability assessments—such as N-1 contingency analysis, Monte Carlo simulation, and scenario tree generation—struggle to capture compound risks driven by weather-induced uncertainties. To address this gap, we develop a modular, data-driven framework to estimate the quantile-at-risk (or value-at-risk) of a user-defined resilience metric (e.g., unserved energy or outage duration), conditioned on temperature forecasts and system parameters. The framework integrates historical data on generator outages and recovery, a probabilistic scenario generator, and an adaptive stochastic importance sampling (A-SIS) procedure for efficient quantile estimation. Applied to the New England region, our framework identifies gas generator failure patterns that significantly elevate system risk under fuel and demand uncertainties driven by extreme winter conditions. Its modular design enables users to define resilience metrics, incorporate updated data, and interpret risk drivers. This work offers a scalable, interpretable approach to risk-informed energy system planning and operations under increasing climate and infrastructure uncertainty.

3 - Pipeline Capacity Allocation Mechanisms to Mitigate Fuel Supply Interruptions at Gas-Fired Power Plants

Ashish Radhakrishnan, Pennsylvania State University, STATE COLLEGE, PA, United States, Chiara Lo Prete

Fuel supply interruptions at gas-fired power plants during extreme winter weather remain a persistent challenge to electric system reliability. These disruptions are typically driven by spikes in residential and commercial gas demand, constrained pipeline capacity, and rigid legacy gas curtailment protocols that limit the flexibility to redirect gas to generators that contribute to power system reliability during periods of system stress. Although gas-fired power plants often hold firm transportation contracts, they are treated on equal footing with other firm customers, such as industrial users, without consideration for their critical role in maintaining electric system reliability during winter emergencies. This study investigates the design and analysis of pipeline capacity allocation mechanisms intended to reduce the risk of fuel supply interruptions at power plants during winter emergencies. We evaluate the cost and operational impacts of these mechanisms using a large-scale optimization model of the joint operation of natural gas and electricity systems. The model incorporates competing gas demand across sectors (residential, commercial, industrial and power plant), gas contract types, and emergency gas curtailment priority protocols. The proposed mechanisms for pipeline capacity allocation are tested on a case study representing the Northeastern U.S. during the 2014 Polar Vortex, a period characterized by extreme cold and significant stress on natural gas and electricity infrastructure. The analysis

provides insights into how alternative pipeline capacity allocation mechanisms could potentially improve gas-electric system coordination and enhance the electric system reliability under severe winter conditions.

4 - Steady Flow of Natural Gas in Pipeline Networks - Modeling Choices and Their Implication in Estimating Impact of Extreme Events

Kaarthik Sundar, Los Alamos National Laboratory, Los Alamos, NM, United States, Shriram Srinivasan, Anatoly Zlotnik, Russell Bent

Steady flow of natural gas through interstate pipeline networks is governed, in the most general sense, by nonlinear differential equations. To make these equations more tractable, it is common to neglect several mathematically inconvenient terms and thereby reduce them to a system of nonlinear algebraic equations. Typical simplifying assumptions include: (1) treating natural gas as an ideal gas (to govern the pressure–density relationship), (2) assuming an isothermal (constant-temperature) regime throughout the entire network, and (3) neglecting elevation changes. These simplified equations are widely used for operational optimization, state estimation, and assessing the impact of extreme events on gas-fired generators that, in turn, feed into the power grid. However, the pressure and flow estimates produced by such models can differ substantially from actual values in the pipeline. In this talk, we will begin with first principles, present the governing equations, and then detail the common simplifications found in the literature. Using simple pipeline models, we will demonstrate the errors introduced by each approximation and provide guidance on when each simplification is—or is not—appropriate for different variants of the steady-flow problem. Our goal is to help attendees select the modeling approach best suited to their specific application.

SD30

Building B Level 2 B204

Recent Advances in Power Systems Optimization

Invited Session

ENRE: Electricity

Chair: Anirudh Subramanyam, The Pennsylvania State University, University Park, PA, 16802, United States

1 - Synthesizing Grid Data with Cyber Resilience and Privacy Guarantees

Vladimir Dvorkin, University of Michigan, Ann Arbor, MI, United States, Shengyang Wu

Differential privacy (DP) provides a principled approach to synthesizing data (e.g., loads) from real-world power systems while limiting the exposure of sensitive information. However, adversaries may exploit synthetic data to calibrate cyberattacks on the source grids. To control these risks, we propose new DP algorithms for synthesizing data that provide the source grids with both cyber resilience and privacy guarantees. The algorithms incorporate both normal operation and attack optimization models to balance the fidelity of synthesized data and cyber resilience. The resulting post-processing optimization is reformulated as a robust optimization problem, which is compatible with the exponential mechanism of DP to moderate its computational burden.

2 - From Static to Dynamic: Rethinking Transmission Limits in Wholesale Electricity Markets

Yury Dvorkin, Johns Hopkins University, Baltimore, MD, United States

Extreme weather events increasingly strain power system reliability, exposing limitations in traditional transmission planning and operations. Static line ratings, based on worst-case and conservative assumptions, often underutilize available capacity. While Dynamic Line Ratings (DLRs) offer a promising alternative by reflecting real-time conditions, their integration into market operations remains limited. This talk introduces a stochastic market framework that treats DLR-enabled transmission capacity as a time-dependent, stock-like resource. Using a conservative approximation of line temperature evolution, we formulate a chance-constrained multi-period DC optimal power flow problem and derive a convex reformulation of its non-convex constraints. This approach enables the calculation of energy and reserve prices while ensuring reliability under uncertainty. We demonstrate the model's efficacy and scalability on both an 11-zone and a full 1814-bus NYISO system, highlighting how DLRs can be operationalized for market-based dispatch and pricing under extreme conditions.

3 - Optimization Proxies Using Limited Labeled Data and Training TIME &Ndash; a Semi-Supervised Bayesian Neural Network Approach

Deepjyoti Deka, MIT, Cambridge, MA, United States, Parikshit Pareek, Kaarthik Sundar, Sidhant Misra, Abhijith Jayakumar

Constrained optimization problems arise in various engineering systems such as inventory management and power grids. Standard deep neural network (DNN) based machine learning proxies are ineffective in practical settings where labeled data is scarce and training times are limited. We propose a semi-supervised Bayesian Neural Networks (BNNs) based optimization proxy for this complex regime, wherein training commences in a sandwiched fashion, alternating between a supervised learning step for minimizing cost, and an unsupervised learning step for enforcing constraint feasibility. We show that the proposed semi-supervised BNN outperforms DNN architectures on important non-convex constrained optimization problems from energy network operations, achieving up to a tenfold reduction in expected maximum equality gap and halving the inequality gaps. Further, the BNN's ability to provide posterior samples is leveraged to construct practically meaningful probabilistic confidence bounds on performance using a limited validation data, unlike prior methods.

4 - Madnlp.Jl, Madncl.Jl, and Madqp.Jl: Gpu Solvers for Large-Scale Optimization

Alexis Montoison, Argonne National Laboratory, Chicago, IL, United States, François Pacaud, Sungho Shin, Mihai Anitescu

In recent years, the development of scalable continuous optimization solvers on GPUs has made significant progress, primarily driven by advances in GPU-based sparse linear algebra.

This talk presents a suite of Julia-based open-source solvers designed to efficiently tackle large-scale linear, quadratic, and nonlinear programming problems with full GPU acceleration.

We first introduce MadNLP.jl, a GPU-native solver for nonlinear programming (NLP) based on interior-point methods and sparse direct solvers.

It was the first solver in the suite and remains central to solving general nonlinear problems efficiently.

Building on this foundation, MadNCL.jl is a robust meta-solver that orchestrates multiple NLP solves to handle degeneracy, ill-conditioning, and to achieve high-accuracy solutions with tolerances below 1e-8.

The latest solver MadQP.jl, targets large-scale linear and quadratic programs (LP / QP).

All three solvers are based on second-order algorithms, enabling greater robustness and accuracy compared to purely first-order methods such as PDLP.

We present performance results on real-world benchmark instances, demonstrating substantial speedups between CPU and GPU implementations, while maintaining high solution quality.

5 - Robust Power Grid Defense Under Attacker Budget Uncertainty

Andrew Mastin, Lawrence Livermore National Laboratory, Livermore, CA, United States

Defending the power grid from attacks is often modeled in three stages. In the first stage, a defender hardens up to h components, making them invulnerable. In the second stage, an attacker disables up to k of the remaining vulnerable components. Finally, the system operator makes remedial adjustments to optimize performance on the surviving infrastructure. For a fixed attacker budget k , the Best Response Intersection algorithm has demonstrated the ability to identify optimal hardening strategies on large systems. However, experiments show that such strategies are not robust when the attacker's budget differs from the assumed value. We present a modified version of the Best Response Intersection algorithm to produce hardening strategies that are effective across a range of attacker budgets. We demonstrate its effectiveness on standard test problems.

SD31

Building B Level 2 B205

Energy Systems

Contributed Session

Contributed

Chair: Giulia Fede, Norwegian University of Science and Technology (NTNU), Trondheim, N/A, Norway

1 - Optimization of Ship-Deployed Uav Collaborative Operation for Far Offshore Wind Turbine Blade Inspections**Feng Hui, Harbin Engineering University, Harbin, Heilongjiang Province, China, People's Republic of, YuZhen Hu**

Abstract With the far offshore and large-scale development of offshore wind power, inspecting wind turbine blades has become increasingly challenging. Unmanned Aerial Vehicles (UAVs), with flexible, cost-effective, and safe advantages have become an ideal alternative to manual inspection, and the ship-deployed UAV collaboration mode provides a novel solution to overcome endurance limitations and expand inspection coverage. This study focuses on optimizing the ship-deployed UAV collaborative operation for offshore wind turbine blade inspections. It considers the UAV's continuous inspection capabilities and flexible interactions with the ship to plan collaboration routes that minimize the inspection time span. To meet the multi-scale inspection requirements, this study proposes a two-level branch-and-bound algorithm and a two-stage greedy heuristic with slack, both embedding a second-order cone programming model to optimize UAV launch and retrieval locations. Numerical experiments validate the superiority of the collaborative mode and the solution method, and investigate the influence of key UAV parameters on inspection plans across various scenarios, providing decision support for efficient unmanned inspection of offshore wind turbine blades.

Keywords far offshore wind farm; wind turbine blade inspection; ship-deployed UAV; Path Optimization

2 - Integrated Design of a Green Hydrogen Production System Under Climate and Demand Uncertainty**MINZHUO HUANG, King Abdullah University of Science and Technology, Jeddah, Saudi Arabia, Ricardo Lima, Omar Knio**

This study presents a comprehensive modeling framework for designing an integrated green hydrogen production systems in hot, arid regions like Saudi Arabia. These areas offer strong solar potential but face freshwater scarcity and fluctuating renewable energy availability. Furthermore, climate change introduces long-term shifts in resource availability and operational reliability, while hydrogen demand profiles remain uncertain due to varying local and global applications. To address this, the model incorporates CMIP6 climate projections and multiple demand profiles, including a local continuous demand and an export-oriented pattern.

The system integrates photovoltaic and wind generation, concentrated solar power with thermal storage, batteries, seawater reverse osmosis desalination, PEM electrolysis, and hydrogen storage.

A two-stage stochastic program is proposed: the first stage determines investments in energy, water, and hydrogen infrastructure, while the second optimizes operations under hourly weather and demand scenarios to minimize total cost and the leveled cost of hydrogen (LCOH).

To ensure computational tractability over multi-year horizons with hourly resolution, the model is reformulated as a mixed-integer linear program from a mixed-integer nonlinear program. Case study for Saudi Arabia demonstrates system's operational resilience and ability to minimize expected LCOH across varying climate and demand scenarios. Also, detailed results, including the value of stochastic solution and perfect information are discussed, as well as the system designs and LCOH obtained with alternative climate scenarios.

This work underscores the necessity of incorporating uncertainty into green hydrogen infrastructure planning and offers a scalable, technically feasible model that can be adapted to other regions with similar renewable resource profiles.

3 - An Exact Penalty Riemannian Alternating Direction Method of Multipliers**Gino J. Lim, University of Houston, Houston, TX, United States, Farzane Ezzati, Zhijie Sasha Dong**

We consider a class of binary programming problems with convex objectives. Such problems arise in diverse applications, including microgrid planning, portfolio optimization, production planning, and machine learning tasks such as sparse principal component analysis, feature selection, and prototype selection in clustering. We propose an Exact Penalty Riemannian Alternating Direction Method of Multipliers (EPR-ADMM) to solve this class of discrete non-convex problems. The binary constraints are reformulated via an exact penalty term in the objective and a $n - 1$ -dimensional sphere manifold embedded in Euclidean space. Each iteration of the algorithm involves computationally simple steps. We analyze the algorithm's computational complexity and establish convergence to first-order KKT points, achieving an ϵ -tolerance satisfaction of the exact penalty conditions. To the best of our knowledge, this is the first work to investigate exact penalty properties in Riemannian optimization within the ADMM framework. Numerical experiments on representative problem instances demonstrate the effectiveness of the proposed method.

Keywords: Binary Programming, Riemannian Manifold, Exact Penalty, ADMM.

4 - A Post Processing on Locational Marginal Costs Considering Fixed Losses on DC Optimal Power Flow**Bruno Amaral, PSR, Rio de Janeiro, Brazil, Tiago Andrade, Sergio Granville, Luiz Carlos da Costa**

The economic dispatch problem is used by system operators to determine optimal generation and storage schedules that minimize operational costs while satisfying physical, economic, and environmental constraints. A key output of these models is the locational energy price, which varies across both space and time. Accurate network modeling improves price signals by capturing equipment limits and the dynamics of energy flows. In a simplified setting without congestion or explicit losses, neighboring nodes would have identical prices. In practice, however, network losses introduce spatial price differences even in the absence of congestion.

Explicitly modeling losses is computationally demanding for large networks, making such models intractable for large systems or long planning horizons. To address this, we propose an efficient method to approximate the effect of losses on nodal prices. Rather than modeling losses directly, our approach post-processes prices from simulations using a simplified network based on Power Transfer Distribution Factors (PTDFs), with fixed losses incorporated as nodal demands. While fixing the losses yields reasonable generation and flow results, it fails to reflect well in nodal prices.

To overcome this, we apply a post-processing adjustment to marginal prices using PTDFs and the optimal generation response, estimating the impact of losses on prices when marginal demand increases at each node. Our results show that the method accurately captures both spatial and temporal variations in nodal prices while significantly reducing computational requirements. Thus, we are able to solve large systems with complex network with limited computational resources.

5 - Enhancing Power Grid Resilience: a Risk-Averse Stochastic Optimization Approach for Infrastructure Hardening**Minhang Zhou, University of Arizona, Tucson, AZ, United States, Pavlo Krokhmal**

Hurricanes remain a major threat to overhead distribution networks. This study investigates covered conductors as a cost-effective hardening option that reduces weather-related line faults and service interruptions. We formulate a two-stage stochastic program in which the degree of conductor coverage is selected in the first stage, while the second-stage minimizes total load shedding once hurricane damage is realized. A principal challenge is the decision-dependent uncertainty. To address risk preferences, several stochastic optimization models are proposed, including the traditional risk-neutral formulation and risk-averse extensions based on both linear and non-linear

coherent measures of risk. Exact solution algorithms that employ Benders decomposition and polyhedral approximations of nonlinear constraints have been proposed. Computational experiments on the IEEE 33-bus system demonstrate the properties of the constructed models, as well as the efficiency of the proposed algorithms.

6 - Desalination Investment Decisions in Water-Stressed Regions

Yolanda Matamala, Universidad Católica de la Santísima Concepción, Concepción, Chile

Water scarcity represents a critical challenge for sustainable development, especially in water-stressed regions where projections indicate 33 countries will face severe water scarcity by 2040. This study proposes a methodology based on a Stackelberg model to evaluate the synergy between freshwater markets and electricity, considering the integration of desalination technologies and their impact on electricity prices. Our research addresses two critical uncertainties: the risk of increasing water scarcity and the impact on energy systems from incorporating desalination plants. Results demonstrate that integrating desalination with freshwater management creates viable pathways for emission reduction through green and blue hydrogen production, providing insurance against water availability risks while generating a significant increase in electricity demand. Our analysis indicates that balanced technological portfolios incorporating diverse renewable sources, energy storage, and strategic desalination integration offer more resilient pathways to achieve both water security and decarbonization. The methodology provides policymakers with robust frameworks for developing strategies that address the complex interrelationships between energy production, water resources, and uncertainty management in water-stressed contexts.

7 - Optimization-Based Electricity Pricing for Effective Demand Response

Joonmo Sung, Yonsei University, Seoul, Korea, Republic of, Soongeol Kwon

This study proposes a framework for establishing electricity pricing schemes to support the effective implementation of demand response programs. The proposed framework utilizes an optimization model that integrates energy-aware production scheduling, aiming to encourage load shifting through a combination of energy charges and demand charges based on peak demand. To effectively manage the peak demand, it is essential to design a balanced pricing structure between time-varying pricing and demand charge rate. The framework enables electricity utilities to determine appropriate demand charge levels to regulate peak demand by analyzing the interrelationship between energy charges and peak demand. Numerical experiments were conducted using real-world electricity tariffs in a parallel machine scheduling environment. The results demonstrate that the proposed framework aids utilities in making informed pricing decisions under various operational conditions. This study contributes to promoting voluntary changes in electricity consumption patterns and enhancing the success of demand response programs.

8 - Optimization of Hydrogen Supply Operations for Decarbonizing Energy-Intensive Industries: a Multi-Timescale Rolling Horizon Approach

Giulia Fede, NTNU, Trondheim, Norway, Fabio Sgarbossa, Daniel F. Silva

The replacement of natural gas with green hydrogen, produced via water electrolysis using renewable energy sources, is increasingly recognized as a promising decarbonization strategy for energy-intensive industries. However, the feasibility of this transition largely depends on the economic viability of hydrogen supply, highlighting the importance of optimized hydrogen supply operations to minimize costs. This study develops a Mixed-Integer Nonlinear Programming (MINLP) model to optimize on-site green hydrogen production, which can be complemented by external hydrogen delivery when required. The model explicitly captures key operational dynamics, including variable loads, transitions between operational states, and efficiency degradation of the electrolyzer stack. The MINLP formulation is embedded within a multi-timescale rolling horizon framework. The rolling horizon approach enables reactive hydrogen supply scheduling to address uncertainties, such as fluctuations in renewable energy availability and hydrogen demand. The multi-timescale structure distinguishes between short-term decisions (e.g., operational states and variable loads of the electrolyzer) and long-term decisions (e.g., scheduling of truck deliveries), consequently accounting for their distinct implementation lead time requirements. Overall, the proposed approach aims to support cost-effective integration of green hydrogen into energy-intensive industrial processes while ensuring continuous and reliable fuel supply.

SD32

Building B Level 2 B206

Sustainable Operations

Invited Session

ENRE: Environment and Sustainability

Chair: Chunjie Zhao, Clark University, 950 MAIN STREET, Worcester, MA, 01610-1477, United States

1 - Blockchain Driven Scope 3 Emission Disclosure for Eco-Conscious Consumer-Oriented Platform Economy

Yuqian Wang, Renmin University of China, Beijing, China, People's Republic of, Jianghua Wu

Platform-led supply chains are adopting blockchain-powered Scope 3 carbon emissions in addition to Scope1-2 carbon emissions when designing their carbon footprint reduction (CFR) project to accommodate eco-conscious consumers. There are two common types of Scope 3 emissions disclosures: Closed-loop (C-Scope 3) which includes emissions from all members in the supply chain, and Cradle-to-Gate (U-Scope 3) which includes emissions only from upstream members. This paper considers a supply chain consisting of a platform, an upstream supplier and a downstream logistics service provider. We explore the effectiveness of the platform and supply chain in achieving economic and environmental goals under different carbon disclosure. Firstly, our results suggest that when the CFR subsidy is low with lower initial emission transparency and traceability cost, adopting U-Scope 1-3 and C-Scope 1-3 emission disclosure strategies can benefit the platform and the supply chain. However, C-Scope 1-3 (instead of U-Scope 1-3) is more likely to achieve win-win (environment-economic) Pareto improvement for both the platform and the supply chain. Secondly, when traceability cost (CFR subsidy) is low, C-Scope3(Scope1-2) is the most environmentally and economically beneficial strategy. Thirdly, the U-Scope3 (C-Scope3) optimal CFR subsidy is larger when the upstream (downstream) margin is larger. Moreover, as for designing logistic mode, when the cost of self-run logistic is low, lower CFR subsidy and traceability cost leads to adopting self-run logistic strategy benefits the platform and the supply chain. This provides management insights for disclosure strategy, emission incentive design and logistic mode selection.

2 - ESG Reports as Decision Support: Topic Modeling, Investor Behavior, and Market Implications

Yichuan Wang, Northern Arizona University, Flagstaff, AZ, United States, Jun Cheng, Yuyu Liu, Suying Sun

We examine the informational value of corporate ESG reports for sustainable investment and decision-making. Using a large sample of ESG reports from Chinese listed firms, we apply Latent Dirichlet Allocation (LDA) to extract narrative topics and employ an XGBoost model with Bayesian optimization to predict institutional investors' shareholding decisions. Interpretable machine learning methods (e.g., feature importance, SHAP values, and accumulated local effects) reveal which ESG topics and tones most strongly influence investor behavior, uncovering nonlinear effects often overlooked in prior work. In addition, we construct hedging portfolios based on ESG narrative topics and tones, many of which generate significant excess returns. These findings suggest that ESG reports contain underutilized decision-relevant information, with implications for sustainable investment, market efficiency, and corporate ESG disclosure strategies. Our study contributes to the literature by integrating textual analysis with advanced machine learning to demonstrate how ESG narratives shape capital allocation and support sustainable operations.

Keywords: ESG, Sustainable Investment, Textual Analysis, Machine Learning

3 - Motives and Results of Sustainable Supply Chain Practices: An Empirical Investigation from An Integrated Theoretical Perspective

Xianghui Peng, Penn State University - Erie, Erie, PA, United States, Lu Xu, Xinyu Wei, Ying Cao, Victor Prybutok

This study examines the existing approaches and frameworks on sustainable supply chain management (SSCM). Theories and practices from other fields, including quality management, supply chain management, and management (stakeholder theory and institutional theory), are explored to identify their contribution to theory development in SSCM. Drawing upon these theories and practices, a research model for SSCM is developed with an integrated theoretical perspective. The developed research model

integrates leadership, instrumental motives, relational motives, and moral motives, along with SSCM practices to drive economic results, social results, and environmental results of an organization. The developed research model provides a comprehensive overview of the functioning mechanism of SSCM at an organizational level. We empirically test and validate the research model through survey data using a structural equation modeling approach. The construct measurements are tested and validated first. The research model is supported by empirical results, which provide rich evidence for theory development and guiding industry practices.

SD33

Building B Level 2 B207

Decision Making in Healthcare

Invited Session

Health Applications Society

Chair: Yili Wang, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Kuofu Liu, University of Michigan, Ann Arbor, MI, United States

1 - Location and Capacity Planning of Substance use Treatment Facilities for Maximizing Access and Population Health Outcomes

Zixuan Feng, The Pennsylvania State University, University Park, PA, United States, Qiushi Chen, Paul Griffin, Pengyi Shi

The opioid epidemic has devastated communities across the US, causing more than 800,000 deaths over the past two decades and costing \$2.7 trillion economic burden in 2023 alone. As more resources became available for expanding treatment capacities, a practical question arises on how to optimally allocate treatment resources geographically. Although travel distance has been shown to significantly impact treatment retention, few studies have considered this critical factor for strategic planning of treatment capacities. To address this gap, we developed a data-driven approach to guide the capacity allocation aimed at curbing the impact of the opioid epidemic. We formulated a discrete-time compartment model to characterize the dynamic course of opioid use disorder, in-treatment, and recovery for patients from different regions and treatment facilities at different locations. In this model, patients' travel distance affected not only treatment retention rates but also their preferences in choosing a treatment location when initiating treatment. We further incorporated this dynamic model in a mixed-integer program that optimizes the allocation of treatment capacity to possible treatment facility locations to minimize the overdose deaths. In our computational study, we estimated treatment retention rates based on travel distances using patient-level data from a clinical study, and calibrated other transition parameters against public health data for the opioid epidemic at the state level. We compared total overdose deaths over the planning horizon under the optimal policy derived from our proposed approach with other heuristic capacity allocation strategies.

2 - A Multi-Agent Reinforcement Learning Algorithm for Personalized Recommendations in Bipolar Disorder

Sidian Lin, Harvard University, Cambridge, MA, United States, Soroush Saghaian, Jessica M. Lipschitz, Katherine E. Burdick

This study introduces a novel multi-agent reinforcement learning (MARL) algorithm designed for identifying and optimizing personalized recommendations in bipolar disorder. The algorithm leverages longitudinal offline data from wearables to recommend self-care strategies tailored to individual patients. We focus on self-care strategies involving physical activity (measured by steps), sleep duration, and bedtime consistency, aiming to reduce the periods of mood exacerbations. A key innovation of our MARL approach is the integration of copulas to model inter-agent dependencies, enhancing coordination among agents and improving policy learning. Findings suggest that following our algorithm's self-care recommendations could significantly reduce periods of elevated mood symptoms, resulting in improved overall well-being. Finally, the algorithm offers important clinical insights for treating bipolar patients, and shows promising theoretical properties independent of the specific application. Thus, this work not only advances MARL applications in personalized healthcare but also provides a new algorithmic approach for adaptive interventions in a wide range of chronic diseases.

3 - Dynamic Programming for Coordinated Care Across Multiple Chronic Conditions

Kuofu Liu, University of Michigan, Ann Arbor, MI, United States, Luke DeRoos, Amirhossein Moosavi, Mariel Lavieri, Jason Miller, Joshua Stein

In the US, four in ten adults live with two or more chronic health conditions. For many of these conditions, treatment timing is determined by the maximum safe treatment interval (the longest allowable gap between consecutive treatments). Coordinating care across multiple conditions can reduce patient costs and improve their experience without compromising health outcomes. In this study, we develop a dynamic programming approach to optimize treatment synchronization of patients with multiple chronic conditions based on their maximum safe treatment intervals. Our analysis determines when it is optimal to synchronize treatments across conditions versus treating them independently, and we derive the corresponding optimal synchronization policy. For complex scenarios involving many conditions, we propose a closed-form heuristic and establish a bound on its regret. We demonstrate the approach through a case study of patients with bilateral (two-eye) age-related macular degeneration, a condition in which treatment needs often differ between eyes. The results indicate that synchronizing treatments could save US patients up to \$582 million annually in direct costs.

4 - Dynamic Nonuse Risk Estimation for Timely Interventions in Transplantation

Umur Kose, Virginia Tech, Blacksburg, VA, United States

Over 90,000 patients in the U.S. await kidney transplants, yet nearly 28% of recovered kidneys go unused each year. Existing risk estimation models fall short in predicting kidney nonuse because they often overlook key operational realities of the allocation process including the dynamic accumulation of cold ischemia time (CIT), evolving responses from transplant centers during the match run, and critical geographic and logistical factors. Moreover, these models treat nonuse as a binary outcome, ignoring its inherently time-dependent nature. To address these limitations, we develop a dynamic survival analysis framework to predict kidney utilization outcomes in real time during the match run. Our approach combines baseline donor characteristics, such as demographics and clinical lab values, with temporal and spatial features, including accumulated CIT, rejection reasons, the proximity of donor hospitals to major airports, and the density of nearby transplant centers. Our models achieve strong performance as measured by the time-dependent concordance index. Feature importance analysis using SHAP values highlights that several dynamic and spatial variables play a dominant role in predicting kidney nonuse. Our framework can be operationalized within organ procurement workflows to identify kidneys at elevated risk of nonuse and trigger timely, targeted interventions before CIT accumulation reduces transplant viability.

5 - Modeling the Potential Effects of Obstetric Unit Closures on Delivery Volume and Distance to Care

Lauren Steimle, Georgia Institute of Technology, Atlanta, GA, United States, Meghan Meredith, Margaret Carrel, Stephanie Radke

There are known risks associated with childbirth in rural obstetric units with low delivery volume. However, there are concerns that the consolidation of patient volume through the closure of rural obstetric units may reduce already limited access to obstetric care. We use a mathematical optimization model to determine strategic obstetric facility closures that would ensure all open obstetric units have at least 500 deliveries per year, while minimizing the overall travel distance to obstetric care. The model assumes all patients seek care at their closest obstetric care facility and uses data from the U.S. Census Bureau for two states: Iowa and Georgia. In the model, Georgia currently has 15 low-volume (<500 annual births) rural hospitals. After 10 strategic closures, all obstetric units are projected to have at least 500 annual births. Among those who would have to travel farther due to the closures, the average travel distance would increase by 10.6 miles (from 8.7 to 19.3 miles). Iowa has 34 low-volume rural hospitals. After 25 closures, all obstetric units are projected to have at least 500 annual births. The average travel distance would increase by 14.5 miles (from 11.5 to 26.0 miles) among those who would have to travel farther. This conceptual model demonstrates the trade-offs between the consolidation of delivery volume and geographic access to care. This work illustrates that coordinated and thoughtful closures of low-volume obstetric units that are not geographically isolated may result in the consolidation of delivery volume with small impacts on travel distance to care.

SD34

Building B Level 2 B208

Global Optimization Methods in Robust and Stochastic Programming

Invited Session

OPT: Global Optimization

Chair: Shixuan Zhang, Texas A&M University, College Station, TX, United States

1 - Global Decomposition Methods for Stochastic Minlp

Joseph Scott, Georgia Institute of Technology, Atlanta, GA, United States, Jingzhi Yang

This talk presents results from a numerical study identifying key bottlenecks in decomposition algorithms for solving nonconvex two-stage stochastic mixed-integer nonlinear programs (SMINLPs). SMINLPs are difficult to solve due to nonconvexity and large scale. To address this, decomposition methods like those by Cao & Zavala (CZ) and Li & Grossmann (LG) project the problem into the first-stage space and perform Branch-and-Bound (B&B) on the first stage only while still guaranteeing global optimality. Good performance of both methods is only observed when the problem has a small first-stage decision space or the problem is solved at the root node using cuts or bound tightening. The slow convergence of both methods in other problems is hypothesized to be caused by the cluster problem—a very large number of small B&B nodes in the immediate vicinity of the global solution.

From clustering theory, first-order convergence lower bounding scheme suffice to eliminate clustering at a constrained minimizer if the objective grows linearly in all feasible directions. Our numerical study on five benchmark SMINLPs showed that both methods are at least first-order convergence empirically confirmed by the results from prior work by Robertson et al. who investigated the convergence order of CZ and LG methods under different assumptions about problem geometry. However, two SMINLPs still show signs of clustering in contrast to the prediction by the clustering theory. The talk will conclude with a discussion of possible paths forward given these new observations.

2 - Moment Relaxations for Wasserstein Distributionally Robust Optimization

Shixuan Zhang, Texas A&M University, College Station, TX, United States, Suhan Zhong

We study moment relaxations for data-driven Wasserstein distributionally robust optimization. In particular, we identify conditions that ensure asymptotic consistency of the relaxations, for both single- and two-stage problems, as the Wasserstein radius approaches zero. Numerical experiments are also included to show the effectiveness of these relaxations.

3 - Disjunctive Benders Decomposition

Geunyeong Byeon, Arizona State University, Tempe, AZ, United States, Kaiwen Fang, Inho Sin

This paper presents a novel approach to enhancing Benders decomposition by incorporating the generation of valid inequalities for the convex hull of the Benders reformulation, addressing the limitation of conventional Benders decomposition, which typically generates cuts that are only tight with respect to the continuous relaxation. Our method efficiently integrates disjunctive programming theory with Benders decomposition, introducing a new routine that leverages existing cut-generating oracles for uncovering constraints required to construct valid inequalities for the convex hull. For mixed-binary linear programs, this approach eliminates the need to solve the master problem as a mixed-integer program entirely. Additionally, we extend the a posteriori strengthening and lifting procedure for lift-and-project cuts into the Benders decomposition framework in a straightforward and efficient manner, and present an approximate routine for generating lift-and-project cuts. Numerical studies on large-scale uncapacitated and capacitated facility location problems demonstrate the effectiveness of our approach, reducing the number of nodes explored in the branch-and-bound tree by orders of magnitude to reach the lower bound attained by a conventional Benders decomposition.

4 - Multistage Robust Optimization with Predictions

Shuai Li, Georgia Institute of Technology, Atlanta, GA, United States, Xin Chen, Weijun Xie

We study a single-dimensional, multi-stage robust control problem with linear system dynamics, convex state and action costs, and box-type uncertainty sets. Beyond the standard robust control setup, we incorporate a key feature: the uncertainty sets shrink as the decision-maker approaches each stage, reflecting the progressive acquisition of information. The degree of shrinkage is assumed to be known in advance. This model extends the classical robust control framework and aligns with emerging work on algorithms with predictions.

While dynamic programming is effective in the static uncertainty setting, we show it becomes computationally intractable in our model due to the curse of dimensionality. We therefore focus on the decision rule approach, particularly affine policies. We prove that affine policies remain optimal when uncertainty sets begin to shrink only one step before their realization. However, if the shrinkage begins two steps ahead, affine policies become suboptimal, and we construct a counterexample to demonstrate this.

To evaluate the performance gap of affine policies in this more general setting, we propose two lower bounds: one based on a dual affine policy, which offers tighter bounds but at greater computational cost, and another using information relaxation, which is more scalable but looser. Finally, we validate our theoretical insights through numerical experiments.

5 - Bayesian Optimization of Risk with Student-T Process Prior

Alexander Vinel, Auburn University, Auburn, AL, United States, Armin Khayyer

We consider the problem of optimization of tail risk measures (particularly Conditional-Value-at-Risk, CVaR) for black-box functions. We follow the Bayesian optimization framework with a few variations compared to the traditional implementations. We consider the potential to employ the Student-t process instead of the commonly used Gaussian process as the prior in the Bayesian optimization framework, since it places a larger emphasis on the extreme outliers. To this end, we provide a closed-form formulas for a collection of acquisition functions and then demonstrate how the Student-t process version of Bayesian optimization of risk can be organized for optimization of risk. We will then present some preliminary results with benchmark functions using CVaR as the objective.

SD35

Building B Level 2 B209

AI and OR for Equitable Access to Healthcare Services.

Invited Session

Public Sector OR

Chair: Holly Wiberg, Carnegie Mellon University, Pittsburgh, PA, 15213, United States

1 - Community Factors Influencing the Duration of Buprenorphine Treatment in Individuals with Opioid use Disorder

Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States, Paula Jaimes-Buitron, Kehe Zhang, Wanru Guo, Yanmin Gong, Yuanxiang Guo, Cici Baurer

Opioid Use Disorder is a significant public health issue in the United States, with fewer than 25% of diagnosed individuals starting treatment, and more than half discontinuing within a year. Medications for Opioid Use Disorder, such as buprenorphine, methadone, and naltrexone, are effective in reducing the risk of overdose; however, long-term adherence to treatment is crucial to minimize relapse and mortality rates. This study explores the relationship between the duration of buprenorphine treatment and social determinants of health using data from the IQVIA PharMetrics database spanning from 2006 to 2023. Key factors analyzed include the availability of buprenorphine-waivered providers, the density of mental health services, and socioeconomic vulnerabilities. Buprenorphine treatment episodes were categorized into four duration groups: 0–3 months, 3–6 months, 6–12 months, and over 12 months. A stepwise multinomial regression model was utilized to assess the association between demographic and community-level factors and treatment duration.

The results indicate that patients in areas with lower Social Vulnerability Index scores tended to remain in treatment for longer periods. A greater number of buprenorphine-waivered providers and mental health services were linked to better treatment retention. However, patients enrolled in Medicaid experienced shorter treatment durations, which may be attributed to issues related to provider availability or prior authorization. Additionally, female patients had shorter treatment durations compared to males, likely influenced by caregiving responsibilities and the stigma surrounding OUD. These findings highlight disparities in OUD treatment retention and underscore the need for targeted, evidence-based strategies to address health inequities.

2 - Operationalizing Advanced Primary Care in Employer-Sponsored Insurance

Katherine Bobroske, Morgan Health, Washington, DC, United States, Jacob Caswell, Rowana Ahmed, Joseph Real

In the United States, advanced primary care (APC) is a care delivery model that strives to connect patients to comprehensive care in the primary care setting. Programs provide patients with more direct and integrated access to primary care clinicians, virtual appointments, nutrition and behavioral health counseling, and chronic condition management. The model aims to manage patients preventatively and reduce the patient demand for high cost specialists and avoidable acute/emergency care. While there is evidence that APC effectively improves quality and reduces barriers to care in some Medicare settings, there are unique barriers to overcome when deploying these models in the employer-sponsored insurance (ESI) context.

To better understand if and how APC can improve patient management in ESI, we clustered patients based on their primary care usage, including: primary care visit counts, medical and prescription costs, chronic condition management status, self-reported behavioral health and wellness scores. The algorithm identified 8 distinct clinical personas - including emergency room "frequent fliers," patients disconnected from primary care who struggled with condition management, and healthy patients with above-average wellness scores. We then leveraged a decision tree methodology to develop heuristics that providers could use to classify patients into the clinical personas. The findings provide insight into how APC programs can target their engagement strategies and initiatives to improve the overall impact that APC makes in ESI.

3 - Real-Time Patient Placement in Observation Units: a Capacity-Aware Optimization Approach with Machine Learning

Vera Tilson, University of Rochester, Rochester, NY, United States, Dessislava Pachamanova, Keely Dwyer Matzky

Observation Units (OUs) are short-term care settings designed for patients who require additional evaluation following an Emergency Department (ED) visit. Deciding whether to admit a patient to the OU or to an inpatient ward carries important consequences for clinical outcomes, operational efficiency, and financial costs. Our approach incorporates machine learning to predict which patients are most likely to benefit from OU placement, further we propose a real-time, capacity-sensitive optimization framework to support these critical placement decisions. Using hospital data and queueing theory, we assess the algorithm's performance across various capacity scenarios, highlighting key trade-offs between ED crowding and total hospital length of stay.

4 - Nurse Decision Making During Triage and the Value of Algorithmic Intervention

Anna Zink, Tufts University, Medford, MA, United States

The health care industry is investing billions of dollars on sophisticated algorithms to enhance clinical decision making, but the return on investment is far from guaranteed. How can healthcare organizations make more informed adoption decisions? We study this question in the context of Emergency Department (ED) triage, where triage plays a crucial role in ensuring that patients with urgent needs are treated quickly. Using data from more than 2 million emergency department encounters in 52 clinical sites, we find that nurses misclassify urgent cases; however, when the emergency department is busier and triage mistakes carry higher stakes because of longer waiting times, nurses tend to do a better job of identifying urgent cases. Hispanic, Medicaid, and uninsured patients are systematically undertriaged compared to other patients, and this disparity increases when the ED is busy, despite improved triage accuracy for the average patient. Finally, we demonstrate that implementing an algorithmic solution trained on nurse triage scores alone would perpetuate these biases. Instead, our results find that a more optimal approach is a hybrid one that combines human judgment with algorithmic input to improve triage accuracy and equity.

SD36

Building B Level 2 B210

Advances in Nonlinear Optimization I

Invited Session

OPT: Nonlinear Optimization

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

1 - Random Subspace Methods for Solving Nonconvex Qps

Yunlang Zhu, Lehigh University, Bethlehem, PA, United States, Frank E. Curtis, Aida Khajavirad

Various real-world applications involve the formulation of quadratic programming (QP) problems, and an important area of research is finding ways to find good approximate solutions of large-scale QPs efficiently. Recently, a series of papers have been published about applying random projections to reduce the dimension of the decision variables of QP problems, leading to projected problems of smaller dimension. Motivated by these results, we consider combining certain deterministic components extracted from the original QP problem with random projections so that we can select the subspace to be projected onto based on problem information, not solely randomness. Numerical experiments show that our method yields better approximate solutions than purely random projections with comparable efficiency. Moreover, we provide theoretical guarantees for our method, which shows improved bounds compared to random projections.

2 - Proximal Gradient Methods for Regularized Optimization with Constraints

Xiaoyi Qu, Lehigh University, Bethlehem, PA 18015-3728, PA, United States

The proximal gradient method is a well-established technique for solving regularized optimization problems. However, its extension to problems involving general nonlinear constraints remains largely unexplored. In this talk, we introduce a proximal gradient method specifically designed for such constrained problems. Our approach begins by employing a decomposition strategy to compute trial steps, followed by the use of a merit function to assess the acceptance or rejection of the trial step. Under common constraint qualifications, we establish both worst-case complexity bounds, and subsequence convergence guarantees. Computational results will be presented to demonstrate the effectiveness and robustness of the proposed method.

3 - An Adaptive and Parameter-Free Nesterov's Accelerated Gradient Method for Convex Optimization

Jaewook J. Suh, Rice University, Houston, TX, United States, Shiqian Ma

We propose AdaNAG, an adaptive accelerated gradient method based on Nesterov's accelerated gradient method. AdaNAG is line-search-free, parameter-free, and achieves the accelerated convergence rates $f_k - f_\star = O(1/k^2)$ and $\min_{i \in \{1, \dots, k\}} \|\nabla f(x_i)\|^2 = O(1/k^3)$ for L -smooth convex function f . We provide a Lyapunov analysis for the convergence proof of AdaNAG, which additionally enables us to propose a novel adaptive gradient descent (GD) method, AdaGD. AdaGD achieves the non-

ergodic convergence rate $f_k - f_\star = O(1/k)$, like the original GD. The analysis of AdaGD also motivated us to propose a generalized AdaNAG that includes practically useful variants of AdaNAG. Numerical results demonstrate that our methods outperform some other recent adaptive methods for representative applications.

4 - Nearly Optimal Lp Risk Minimization

Zhichao Jia, Georgia Institute of Technology, Atlanta, GA, United States, Guanghui Lan, Zhe Zhang

Convex risk measures play a foundational role in the area of stochastic optimization. However, in contrast to risk neutral models, their applications are still limited due to the lack of efficient solution methods. In particular, the mean Lp semi-deviation is a classic risk minimization model, but its solution is highly challenging due to the composition of concave-convex functions and the lack of uniform Lipschitz continuity. In this paper, we discuss some progresses on the design of efficient algorithms for Lp risk minimization, including a novel lifting reformulation to handle the concave-convex composition, and a new stochastic approximation method to handle the non-Lipschitz continuity. We establish an upper bound on the sample complexity associated with this approach and show that this bound is not improvable for Lp risk minimization in general through the construction of a nearly matching lower complexity bound.

5 - Popov Mirror-Prox Method for Variational Inequalities

Abhishek Chakraborty, ASU, Tempe, AZ, United States, Angelia Nedich

This paper examines the convergence properties of the Popov mirror-prox algorithm for solving stochastic and deterministic variational inequalities (VIs) under a polynomial growth condition on the mapping variation. Unlike existing methods that require prior knowledge of problem-specific parameters, we propose step-size schemes that are entirely parameter-free in both constant and diminishing forms. For stochastic and deterministic monotone VIs, we establish optimal convergence rates in terms of the dual gap function over a bounded constraint set. Additionally, for deterministic VIs with H^ν-order continuous mapping, we prove convergence in terms of the residual function without requiring a bounded set or monotone mapping, provided a Minty solution exists. This allows our method to address certain classes of non-monotone VIs. However, knowledge of the H^ν-order exponent is necessary to achieve the best convergence rates in this case. By extending mirror-prox techniques to mappings with arbitrary polynomial growth, our work bridges an existing gap in the literature. We validate our theoretical findings with empirical results on matrix games, piecewise quadratic functions, and image classification tasks using ResNet-18, demonstrating the effectiveness of our method.

SD37

Building B Level 2 B211

Sequential Decision-making Under Uncertainty

Invited Session

OPT: Optimization Under Uncertainty

Chair: Harsha Gangammanavar, Southern Methodist University, Dallas, TX, United States

1 - Inverse Optimization In Finite-State Semi-Markov Decision Processes

Nhi Nguyen, University of Minnesota, Minneapolis, MN, United States, Archis Ghatge

Inverse optimization involves finding values of problem parameters that would render given values of decision variables optimal. This is in contrast with the usual forward optimization where decision variables are computed using given values of problem parameters. In the context of finite-state Markov decision processes (MDPs), the inverse optimization literature has focused on imputing two types of inputs that would make a given policy optimal: (i) cost parameters or (ii) transition probabilities. All existing work only considers the case of discrete-time MDPs. We will present an extension to the case of semi-Markov decision processes (SMDPs), with continuous-time MDPs (CTMDP) as a special case. We will utilize Bellman's equations for SMDPs to formulate a parametric version of the inverse problem. In this parametric version, estimates of parameters that characterize costs are given along with a target policy. The goal is to find parameter values that are as close as possible to these estimates and render the given policy optimal. We will present a tractable linear or convex formulation of this problem. Computational experiments on applications will be discussed.

2 - A Gpu Acceleration Procedure in Decomposition Algorithms for Stochastic Linear Programs

Yihang Zhang, University of Southern California, Los Angeles, CA, United States, Suvrajeet Sen

We present an acceleration method based on off-loading the dual bounding ("argmax") scheme by using a GPU to conduct a particular matrix-matrix multiplication which leverages the massive parallelism of GPUs.

3 - Comparative Study of Sampling-Based Multistage Stochastic Linear Programming Algorithms

Zhiyuan Zhang, Southern Methodist University, Dallas, TX, United States, Harsha Gangammanavar

Multistage stochastic linear programming (MSLP) provides a powerful framework for decision-making under uncertainty over time. In this study, we compare two prominent approaches for solving stagewise independent MSLP problems: stochastic dual dynamic programming (SDDP) and stochastic dynamic linear programming (SDLP). We also present a regularized variant of SDDP that, like SDLP, employs basic feasible policies to identify the prox-center. We evaluate the algorithms on benchmark instances of varying size and structure, investigating their convergence behavior and computational efficiency. A series of numerical experiments is conducted to demonstrate the trade-offs among these methods in terms of accuracy and scalability.

4 - A Stochastic Decomposition Method for Multistage Distributionally Robust Optimization Under Streaming Data

Tian Xia, The Ohio State University, Columbus, OH, United States, Harsha Gangammanavar, Guzin Bayraktan

Multistage Stochastic Programming (MSP) is a fundamental framework for modeling sequential decision-making under uncertainty. However, challenges such as curse of dimensionality and random variable's unknown probability distribution make MSP difficult to solve in practice. In this talk, under the streaming data setting, we present a Multistage Distributionally Robust Optimization (MDRO) framework to approximate the original MSP, where the radius of the ambiguity set of distributions shrink as more data is obtained. We propose a Stochastic Decomposition (SD) method specifically tailored to MDRO under streaming data, i.e., an online environment, named as MDRSD. Within the streaming data-driven paradigm, we construct the ambiguity sets using the Wasserstein distance with shrinking radii and develop efficient approximations for both value functions and solutions to MSP. We show conditions under which the sequence of incumbent solutions generated by MDRSD converge asymptotically to those of the original MSP. As an additional theoretical result, we establish that the optimal value of the MDRO converges asymptotically to that of the MSP under our problem setting. Finally in numerical experiments, we validate the efficiency of MDRSD in solving real-world applications under streaming data.

SD38

Building B Level 2 B212

Optimization Society Award Session II

Award Session

Optimization

Chair: Jeff Linderoth, University of Wisconsin-Madison, Madison, WI, United States

1 - Young Researchers Prize

Benjamin Grimmer, Johns Hopkins University, Baltimore, MD, United States

Benjamin Grimmer is awarded the 2025 INFORMS Optimization Society Young Researchers Prize for his paper "Provably Faster Gradient Descent via Long Steps", SIAM Journal on Optimization 34.3 (2024): 2588-2608. For decades, it was unknown whether stepsizes longer than those guaranteeing per-iteration descent could improve the convergence of gradient descent for smooth convex optimization. The cited paper resolves this long-standing open question with a breakthrough result:

carefully designed cyclic schedules that periodically include long steps provably achieve strictly better worst-case guarantees than the classical constant-stepsize method, even though individual iterates may temporarily worsen the objective. Methodologically, the work departs sharply from traditional one-step induction analyses by leveraging a computer-assisted performance-estimation framework. It introduces a “straightforwardness” property for stepsize schedules that can be certified via a semidefinite program, enabling automated discovery and proof of effective long-step patterns. In contrast to its more common role of confirming or refining existing analyses, here, performance estimation enabled the derivation of an unexpected result. The paper has already drawn broad interest and inspired substantial follow-up research, establishing a new foundation for our understanding of gradient descent

2 - Accessible Theoretical Complexity of the Restarted Primal-Dual Hybrid Gradient Method for Linear Programs with Unique Optima

This paper develops the first accessible iteration bound for the restarted primal-dual hybrid gradient method (rPDHG), a method that has emerged as a particularly successful alternative for solving large-scale linear programs. The bound is expressed in closed form using quantities tied directly to the optimal solution and basis, making it both theoretically rigorous and practically computable. The analysis sheds new light on the method’s two-stage performance, sensitivity to perturbations, and practical heuristics, and is further supported by numerical experiments. This single-authored student paper stands out for its accessible analysis, important results, and current relevance, making it a deserving recipient of this prize.

3 - A Simple Uniformly Optimal Method without Line Search for Convex Optimization.

Tianjiao Li, Massachusetts Institute of Technology, Cambridge, MA, United States

Line search (or backtracking) procedures have been widely employed into first-order methods for solving convex optimization problems, especially those with unknown problem parameters (e.g., Lipschitz constant). In this paper, we show that line search is superfluous in attaining the optimal rate of convergence for solving a convex optimization problem whose parameters are not given a priori. In particular, we present a novel accelerated gradient descent type algorithm called auto-conditioned fast gradient method (AC-FGM) that can achieve an optimal rate of convergence for smooth convex optimization without requiring the estimate of a global Lipschitz constant or the employment of line search procedures. We then extend AC-FGM to solve convex optimization problems with Holder continuous gradients and show that it automatically achieves the optimal rates of convergence uniformly for all problem classes with the desired accuracy of the solution as the only input. Finally, we report some encouraging numerical results that demonstrate the advantages of AC-FGM over the previously developed parameter-free methods for convex optimization.

SD39

Building B Level 2 B213

Discrete Optimization for Social Good

Invited Session

OPT: Integer and Discrete Optimization

Chair: Sean Lo, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Online Assignment-Routing for Organ Transplants

Sean Lo, Massachusetts Institute of Technology, Cambridge, MA, United States, Alexandre Jacquillat

TransMedics is a US-based healthcare startup which operate devices for organ transplants. These portable devices preserve the shelf-life of donated organs outside the donor's body, allowing for less time-dependent organ-donor matchings and improved healthcare outcomes. We help TransMedics solve a complex logistics problem in an online setting with a hybrid assignment/routing structure: we assign surgeons, device specialists, and devices to organ donors, and route them to organ recipients via a fleet of planes. We propose an online algorithm which takes into account uncertainty in case arrivals, and demonstrate how it outperforms a greedy approach. Real-world experiments show how our approach improves the volume of transplants served.

2 - Meeting Uncertain Demand with Deadlines: Applications in Defense and Disaster Relief

Louis Chen, Naval Postgraduate School, Monterey, CA, United States, Roberto Szechtman, Vince Vanterpool, Chiwei Yan

We formulate and study the problem of allocating resources across a network towards (uncertain) demands with (certain) deadlines. In particular, allocations are made online, and the decision maker receives “satisficing” feedback in between rounds. This problem encompasses a wide range of applications in defense and beyond. More specifically, it provides new dynamic weapon target assignment and disaster relief models. In addition to introducing these models, we present online policies, structural results, and numerical experiments.

3 - Optimizing Wildfire Suppression: a Branch-and-Price-and-Cut and Double Machine Learning Approach

Leonard Boussioux, UW Foster, Harvard, Seattle, WA, United States, Alexandre Jacquillat, Jacob Wachspress

Wildfires present an escalating threat to communities, infrastructure, and ecosystems, demanding sophisticated strategies for effective suppression. This talk introduces an advanced optimization framework—branch-and-price-and-cut—for coordinating wildfire suppression efforts in complex, dynamic environments. Our two-sided set-partition formulation simultaneously models fire spread dynamics and crew routing decisions on time-expanded networks, integrating critical considerations such as travel cost, crew fatigue, fire damage, and suppression difficulty. We demonstrate computational advances through augmented Generalized Upper-Bound (GUB) cutting planes, dual-aware branching rules, and targeted primal heuristics, achieving substantial improvements in both solution quality and computational efficiency. A comprehensive data-driven case study leveraging historical wildfire data and a double machine learning approach highlights the practical impact of our optimization model, enabling precise estimation of suppression effectiveness and improved decision-making under uncertainty. Results reveal that optimized resource allocation significantly enhances suppression outcomes, prioritizing high-risk fires while balancing crew fatigue and travel costs. This research provides robust, scalable solutions vital for addressing the increasing complexity and severity of wildfire events.

4 - Conformal Inverse Optimization for Adherence-Aware Prescriptive Analytics

Bo Lin, University of Toronto, Toronto, ON, Canada, Erick Delage, Timothy Chan

Inverse optimization has been increasingly used to estimate unknown parameters in an optimization model based on decision data. We show that such a point estimation is insufficient in a prescriptive setting where the estimated parameters are used to prescribe new decisions. The prescribed decisions may be low-quality and misaligned with human intuition and thus are unlikely to be adopted. To tackle this challenge, we propose conformal inverse optimization, which seeks to learn an uncertainty set for the unknown parameters and then solve a robust optimization model to prescribe new decisions. Under mild assumptions, we show that the suggested decisions can achieve bounded out-of-sample optimality gaps, as evaluated using both the ground-truth parameters and the decision maker's perception of the unknown parameters. Our method demonstrates strong empirical performance compared to classic inverse optimization.

5 - Optimizing Threshold-Based Fairness

John Hooker, Carnegie Mellon University, Pittsburgh, PA, United States, Ozgun Elci, Peter Zhang

Threshold-based fairness is designed to balance efficiency and equity in a principled way. It maximizes a social welfare function that measures total benefit up to a threshold at which the distribution of benefits becomes unfair. At this point it begins to switch to a Rawlsian maximin or leximax criterion. An interpretable parameter regulates the relative importance of equity and efficiency in the solution. This talk shows how to give this criterion a practical formulation in an optimization model. It illustrates results on facility location and public transport network problems in which one seeks an efficient solution that provides an acceptable service level to everyone.

6 - Optimizing Wildfire Suppression: A Branch-and-Price-and-Cut and Double Machine Learning Approach

Leonard Boussioux, University of Washington, Foster School of Business, Seattle, WA, United States, Alexandre Jacquillat, Jacob Wachspress

Wildfires present an escalating threat to communities, infrastructure, and ecosystems, demanding sophisticated strategies for effective suppression. This talk introduces an advanced optimization framework—branch-and-price-and-cut—for coordinating wildfire suppression efforts in complex, dynamic environments. Our two-sided set-partition formulation simultaneously models fire spread dynamics and crew routing decisions on time-expanded networks, integrating critical considerations such as travel cost, crew fatigue, fire damage, and suppression difficulty.

We demonstrate computational advances through augmented Generalized Upper-Bound (GUB) cutting planes, dual-aware branching rules, and targeted primal heuristics, achieving substantial improvements in both solution quality and computational efficiency. A comprehensive data-driven case study leveraging historical wildfire data and a double machine learning approach highlights the practical impact of our optimization model, enabling precise estimation of suppression effectiveness and improved decision-making under uncertainty. Results reveal that optimized resource allocation significantly enhances suppression outcomes, prioritizing high-risk fires while balancing crew fatigue and travel costs. This research provides robust, scalable solutions vital for addressing the increasing complexity and severity of wildfire events.

SD40

Building B Level 2 B214

Human AI Interaction in Decision-Making

Invited Session

Service Science

Chair: Kejia Hu, University of Oxford Said Business School, Oxford, United Kingdom

Co-Chair: Yuxuan Zhu, Technical University of Munich, München, 80333, Germany

1 - Dynamic Substitution Strategies for Perishable Products in Food Retailing

Ke Sun, McGill University, Montreal, QC, Canada, Yichuan Ding, Mehmet Gumus

We study the dynamic management of perishable inventory in food retail, characterized by uncertain demand, product perishability, and substitution behavior. In collaboration with a major North American grocery retailer, we develop a multi-period stochastic optimization model that jointly determines replenishment and substitution decisions to maximize long-term expected revenue. We begin by characterizing the structure of the optimal policy and show that it takes the form of a state-dependent threshold policy. However, deriving such thresholds becomes computationally intractable due to the curse of dimensionality. To address this challenge, we propose a heuristic approach based on Lagrangian relaxation, which decomposes the original problem into more tractable subproblems while preserving key system dynamics. Using the real transaction data from our industry partner, we benchmark the proposed approach against commonly used heuristics and observe consistent improvements in revenue. Our results highlight the value of incorporating substitution effects and forward-looking decision-making in the management of perishable inventory. This work contributes to the literature on sustainable retail operations and provides implementable insights for practitioners aiming to improve inventory efficiency in perishable product categories.

2 - Productive Delays

Guillaume Lapierre-Berger, McGill University, Montreal, QC, Canada, Juan Camilo Serpa, Yael Grushka-Cockayne

Contractors strive to minimize project delays at all costs. We demonstrate that some delays can actually improve productivity in the long term, as contractors transfer learning from these delays into other projects within their portfolio. We substantiate this hypothesis using task-specific data from thousands of projects. Our findings reveal that contractors who experienced delays in a given task category were less likely to be delayed later across the same task categories across other projects in their portfolio. Surprisingly, such an early delay has a net positive effect on the portfolio, meaning that the learning spillovers outweighed the costs of the delay. This effect is more pronounced when the delayed tasks lie on the project's critical path.

3 - Causal Estimation and Optimization of Retail Staffing Levels

Xin Wang, University of Toronto, Toronto, ON, Canada, Vahid Sarhangian, Liu Sheng, Opher Baron

This study investigates the causal impact of staffing levels on retail sales, while controlling for key factors such as traffic, inventory, and promotional activities. We begin by applying stratified linear models at the aggregate, store, and time levels to estimate the effect of labor deviations on sales fluctuations. However, treatment effects vary across stores, seasons, and years, resulting in small sample sizes within subgroups and unstable estimates under stratification. To address these challenges, we adopt the Double Machine Learning (DML) framework, which enables the estimation of heterogeneous treatment effects using flexible machine learning models. DML enhances robustness by orthogonalizing the treatment and outcome with respect to high-dimensional confounders, thereby reducing bias from confounding and model misspecification. Unlike traditional stratified models, DML accommodates complex, nonlinear relationships and supports valid statistical inference, even when predictive models are imperfect. Building on these causal estimates, we formulate a roster optimization problem to determine the optimal staffing levels for each store and time period, relative to the baseline labor plan. The objective is to maximize net profit, accounting for counterfactual sales fluctuations and operational constraints in labor scheduling. Finally, we examine the performance of the proposed methods using data from a large retail network.

SD41

Building B Level 3 B301

OR Models in Healthcare Delivery

Invited Session

Health Applications Society

Chair: Ankit Bansal, State University of New York, Binghamton, Binghamton, NY, United States

1 - A Two-Stage Robust Optimization Approach for Anesthesiologist Scheduling with Handovers

Abhishrut Sinha, State University of New York, Binghamton, Binghamton, NY, United States, Ankit Bansal, Osman Ozaltin, Michael Russell

The transfer of operating room responsibilities between anesthesiologists—known as handovers—can increase the likelihood of medical errors and lead to negative health outcomes. To minimize these handovers, we propose a two-stage robust optimization model that assigns anesthesiologists to operating rooms while incorporating key operational constraints. We show that the second-stage problem can be modeled as a longest path problem on a layered acyclic graph and consequently, formulated as a linear program. This linear program is then dualized and embedded into the first-stage model, resulting in a single, integrated Mixed-Integer Linear Programming (MILP) formulation. We also introduce valid inequalities for the first stage and demonstrate their strength by showing that they describe the convex hull of the relaxations of the first-stage integer set. Computational experiments based on real hospital data demonstrate the effectiveness of our proposed approach.

2 - Mixed Integer Linear Program for Family Medicine Residency Scheduling with Clinic First Objectives

Mohammad Al Syouf, State University of New York at Binghamton, Binghamton, NY, United States, Ankit Bansal, Jean-Philippe Richard, Bjorn Berg

Family medicine physicians make up the largest share of the primary care provider (PCP) workforce, accounting for 40% of all PCPs. A notable contributor to the shortage of PCPs is the tendency of family medicine residents to pursue subspecialties or careers as hospitalists. This issue has been primarily linked to a training gap within the family medicine residency curriculum. In practice, family medicine physicians work mainly in outpatient primary care clinics—known within the field as continuity clinics—where they focus on long-term care and maintaining continuity in the physician-patient relationship. However, traditional residency training often emphasizes hospital-based experiences, creating a disconnect between training and practice. The Clinic First paradigm is an educational approach developed to close this gap by prioritizing outpatient clinic experience and reinforcing continuity of care. In alignment with this model, we introduce a Mixed Integer Linear Programming (MILP) model designed to generate annual rotation schedules for family medicine residents. This model captures the complex constraints imposed by accreditation standards and program-specific requirements. It explicitly incorporates two key objectives that reflect the Clinic First paradigm: Clinic Time Consistency (CTC) and Level-Loading (LL). To address the computational challenges, we develop multiple heuristics that efficiently produce high-quality feasible solutions that can be used to significantly speed up the solution of the problem with Gurobi. Computational experiments based on data from an existing residency program demonstrate the effectiveness of our approach.

3 - Bilevel Formulation for Share Allocation in Food Distribution Auctions

Hou-An Chen, NC State University, Raleigh, NC, United States, Osman Ozaltin

This work presents a bilevel optimization framework to address share allocation in Feeding America's Choice system, a market-based mechanism for distributing food donations among food banks. The Choice system uses auctions where food banks bid on food bundles using a specialized currency named shares. While the system offers flexibility and responsiveness to food bank preferences, it has been suffering from low participation from small food banks due to limited shares allocated. To address this, we model the interaction between Feeding America (as the leader) and food banks (as followers) within a single auction window. A centralized decision-making model is first introduced to maximize aggregate utility, followed by a bilevel formulation that captures the strategic behavior of food banks under uncertainty and budget constraints. The proposed model aims to enhance equity and efficiency in food allocation by Choice, offering insights for policy refinement and operational improvements in hunger relief logistics.

4 - Cost-Effectiveness of Personalized Policies for Implementing Organ-at-Risk Sparing Adaptive Radiation Therapy in Head and Neck Cancer

Daniel Suarez Aguirre, Georgia Tech, Atlanta, GA, United States, S. Mohammad Hosseinian, Cem Dede, Raul Garcia, Lucas McCullum, Mehdi Hemmati, Aysenur Karagoz, Mohamed Abdallah S.R. , Stephen Y. Lai, Katherine A. Hutcheson, Amy C. Moreno, Kristy K. Brock, Fatemeh Nosrat, Clifton D. Fuller, Andrew Schaefer

Background and Purpose

The principle of adaptive radiation therapy (ART) is to adjust radiation plans in response to anatomical changes during treatment. The purpose of this study was to develop a decision-making model for implementation of personalized ART that balances the costs and clinical benefits of radiation plan adaptations in head and neck cancer (HNC).

Materials and Methods

Using retrospective imaging data from 52 HNC patients, a Markov decision process (MDP) model was developed to determine optimal timing for plan adaptations based on the difference in normal tissue complication probability (Δ NTCP) between planned and delivered doses to organs-at-risk. To capture the trade-off between the costs and benefits of plan adaptations, the end-treatment Δ NTCPs were converted to Quality Adjusted Life Years (QALYs) and then to equivalent monetary values using a willingness-to-pay per QALY parameter.

Results

The optimal policies were derived for 96 combinations of willingness-to-pay per QALY (W) and re-planning cost (RC) and validated using Monte Carlo simulation for two representative scenarios: (1) W = \$200,000, RC = \$1,000; (2) W = \$100,000, RC = \$2,000. In scenario (1), the MDP model's policy reduced the probability of excessive toxicity (Δ NTCP $\geq 5\%$) to zero (from 0.21 without re-planning) at an average cost of \$380 per patient. In scenario (2), it reduced this probability to 0.02 at an average cost of \$520 per patient.

Conclusions

The MDP model's policies outperformed the current fixed-time (one-size-fits-all) approaches in both clinical and financial outcomes in the simulations.

SD42

Building B Level 3 B302

AI & ML Applications in Healthcare

Invited Session

Health Applications Society

Chair: Sogand Soghrati Ghasbeh, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Mark Van Oyen, Industrial & Operations Engr., 1205 Beal Ave, Ann Arbor, MI, 48109, United States

1 - To be added

Amirhossein Moosavi, University of Michigan, Ann Arbor, MI, United States

2 - Clinical Role Model Generation via Active Fine-Tuning of Large Language Models with Constraint-Aware Feedback

Che-Yi Liao, Georgia Institute of Technology, Atlanta, GA, United States, Esmaeil Keyvanshokoo, Gian Garcia

This work presents a novel framework for generating clinical role models by actively fine-tuning large language models (LLMs) using constraint-aware feedback.

3 - A Tree-Based Policy Reinforcement Learning Algorithm in Non-Stationary Environments with Applications in Healthcare

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

Reinforcement Learning (RL) addresses sequential decision-making in dynamic environments, where optimal policies (i.e., sequences of decisions) are identified to balance long-term consequences under uncertainty.

However, the trust in RL is limited in high-stakes areas. Traditional RL methods operate as black boxes and prioritize long-term cumulative returns without considering human intuition or ethics (i.e., lack of interpretability). In high-stakes settings like healthcare, where humans remain the primary decision-makers, trustworthy recommendations must align with human logic. Through conversations with our medical collaborators, we have identified the clinically relevant notions of interpretability: tree-based structures.

Although emerging RL frameworks have incorporated interpretability metrics as constraints, these methods assume environments that do not change over time (i.e., stationary). This assumption rarely holds in practice, so non-adaptive policies may not remain optimal as conditions change. In healthcare, stationary policies may lead to ineffective treatments and undesirable outcomes as patients' biological conditions fluctuate and diseases progress.

To address these challenges, we proposed an adaptive tree-based algorithm. This algorithm ensures that the internal mechanisms will adapt to changes in non-stationary systems, and the resulting policy will remain inherently interpretable.

4 - Data-Driven Decision Making for Hospital Discharge and Home-Based Care

Mark Van Oyen, University of Michigan, Ann Arbor, MI, United States, Sogand Soghrati Gasbeh, Arlen Dean

Improving patient flow and optimizing hospital resource utilization are critical challenges in modern healthcare systems. Predictive analytics offers a powerful approach to support these goals by enabling proactive decision-making for patient transitions. This presentation introduces a data-driven framework for anticipating care transitions, with a focus on hospital-based or post-acute care pathways.

In one application, we develop and validate a Next-Day Discharge (NDD) prediction model using electronic health record data. Our modeling approach first stratifies patients based on their length of stay and then applies targeted classifiers to predict discharges within each group. This framework captures patient heterogeneity and improves predictive performance over standard models. By identifying patients likely to be discharged in advance, hospitals can better plan for capacity, coordinate follow-up care, and reduce discharge delays.

Further, the framework can be extended to identify patients suitable for Hospital-Care-at-Home (HCaH) programs. HCaH has emerged as a promising alternative to inpatient care for select populations, offering equivalent outcomes with improved patient satisfaction and lower costs. Our predictive model supports this transition by identifying stable inpatients who meet safety and clinical eligibility criteria for at-home care.

For both use cases, we evaluate model performance using real-world hospital data and explore operational implications. The presentation discusses the methodological challenges of deploying prediction models in clinical workflows and the potential for machine learning to enhance hospital operations and patient-centered care.

5 - Comparing AI vs Emergency Physician on ED outcome risk assessment at the point of care

Yiye Zhang, Weill Cornell Medicine/Cornell University, New York, NY, United States

tbd

SD43

Building B Level 3 B303

Data-Driven Decision Making in Health Policy and Emergency Systems

Invited Session

Health Applications Society

Chair: Isabelle Rao, University of Toronto, Toronto, ON, M5S3G8, Canada

1 - Modeling Health and Economic Outcomes of Providing Stable Housing to Homeless Adults with Opioid use Disorder

Isabelle Rao, University of Toronto, Toronto, ON, Canada

The number of people experiencing homelessness (PEH) in the United States is increasing, and drug overdose, primarily driven by opioids, is the leading cause of death in this population. Addressing opioid use disorder (OUD) among PEH is a critical public health priority. We develop a continuous time compartmental model to evaluate the health and economic impacts of providing stable housing to PEH with OUD, without requiring entry into OUD treatment. We simulate a cohort of 1,000 PEH with OUD, comparing outcomes under the housing intervention to those under the status quo (no housing provision) to determine the incremental cost-effectiveness ratio (ICER). Over five years, the housing intervention reduces total deaths from 191 to 140, and reduced total overdoses from 554 to 494. The intervention increases lifetime quality-adjusted life years (QALYs) by 3.59 per person at an incremental cost of \$26,200 per QALY gained compared to the status quo. Across a wide range of sensitivity analyses that could make the housing intervention appear less favorable, the ICER remains below \$90,000 per QALY, indicating that the intervention is cost-effective. Our findings show that investing in stable housing for this marginalized population, even with no requirement to enter OUD treatment, would not only save lives and improve health outcomes but would also be cost-effective. Efforts are urgently needed to improve the health of PEH with OUD. It is essential to understand the effectiveness and cost-effectiveness of housing provision for this marginalized population, as housing status is a key social determinant of health.

2 - Correcting Bias in Value-of-Information: Monte Carlo Errors in Simulation ModelsGary Qian, Stanford, Stanford, CA, United States, Fernando Alarid-Escudero², Hawre Jalal, Jeremy Goldhaber-Fiebert

Monte Carlo (MC) error in microsimulation models introduces a one-sided bias in estimates of value-of-information (VOI), such as the expected value of perfect information (EVPI). While this bias can be reduced with more microsimulated samples, the computational cost can be prohibitive. We aim to develop bias correction approaches under a limited computational budget to feasibly improve the accuracy of VOI calculations using microsimulation models.

3 - The Value of Proxy Information in Location Optimization for Rare Events: Application to Major Trauma Response

Rachel Stephenson, University of Toronto, Toronto, ON, Canada, Sheng Liu, Timothy C. Y. Chan

Major traumas are extremely rare; inherent uncertainty and limited data positions the major trauma response base location problem within the *small-data, large-scale* optimization regime. We analyze the value of proxy information in location optimization for rare events (e.g., major trauma). Specifically, we explore the use of mixed-effects (ME) policies which leverage high variance, unbiased historical observations (e.g., major trauma EMS calls) and low variance, biased proxy data (e.g., population data, low acuity trauma calls) in location decisions. We investigate the structural properties and value of ME policies to present insights to decision makers for informed application to location problems.

4 - Distributionally Robust Group Testing with Correlation Information

Aiqi Zhang, Wilfrid Laurier University, Waterloo, ON, Canada, Daniel Zhuoyu Long, Jin QI, Yu SUN

Motivated by the need for more efficient and reliable methods of group testing during widespread infectious outbreaks, this paper introduces a novel operational improvement to the widely-used Dorfman's group testing procedure, where a single test is conducted on the pooled sample, followed by individual testing of positive pools. Our method minimizes a weighted sum of testing volume and misclassifications, predicated on prevalence rates and inter-individual Pearson correlation coefficients. Recognizing the inherent ambiguity in the population-level joint probability distribution of infection statuses that rises from the correlations between individuals, our approach leverages a distributionally robust optimization (DRO) framework to counteract the worst-case probability distribution. In single-cluster cases, where each pair of subjects are equally correlated, we establish uniform group sizes and connect our analysis to Nash equilibrium principles. Larger testing groups are generally favored under high correlation, whereas individual testing becomes optimal under high prevalence. In multi-cluster cases, where the population is formed by several intra-correlated but inter-independent clusters, we highlight the effectiveness of mixed-cluster testing strategies, particularly at lower levels of prevalence and correlation. This is a notable addition to the prevailing view that advocates pooling correlated individuals. For both single- and multi-cluster cases, we develop polynomial-time solutions and offer practical insights and policy implications for optimal pooling strategies. We demonstrate the trade-offs and benefits of our DRO approach through a thorough comparison with stochastic alternatives, and demonstrate the impact of incorporating correlated infections in a case study using a COVID-19 dataset from Hong Kong.

5 - Cost-Effectiveness of a Precision Digital Health Program for Glucose Management in Pediatric Patients with Newly Diagnosed Type 1 Diabetes

Paul Dupenloup, Stanford University, Stanford, CA, United States, Yupeng Chen, Priya Prahalad, Ramesh Johari, Ananta Addala, Dessi Zaharieva, Mindy Lee, David Maahs, David Scheinker, Margaret Brandeau

Introduction: Youth with Type 1 diabetes (T1D) frequently fail to achieve recommended glycemic targets, increasing their risk of complications. We evaluated the cost-effectiveness of the Teamwork, Targets, Technology, and Tight Range (4T) care model compared to self-monitoring of blood glucose (SMBG) and continuous glucose monitoring (CGM) alone.

Methods: We developed a decision-analytic Markov model simulating 5-year-old newly diagnosed T1D patients over a 20-year time horizon. The model incorporated diabetes-related complications, mortality, and healthcare costs. The 4T intervention combined early CGM initiation, remote patient monitoring via the TIDE platform, and team-based care with standardized glycemic targets.

Results: The 4T model increased quality-adjusted life years (QALYs) compared to both SMBG and CGM alone. Additionally, 4T reduced healthcare costs per patient compared to both alternatives, primarily through reductions in complication-related costs. These findings remained robust across sensitivity and scenario analyses.

Conclusions: The 4T care model simultaneously improves clinical outcomes and reduces costs in pediatric T1D management. National implementation for newly diagnosed youth could yield substantial healthcare savings while enhancing care quality and addressing healthcare disparities. These findings support broader adoption of structured, technology-enabled care models for pediatric T1D management.

SD44

Building B Level 3 B304

Empirical Healthcare Operations

Invited Session

MSOM: Healthcare

Chair: Harshita Kajaria-Montag, Indiana University, Bloomington, IN, United States

Co-Chair: Agni Orfanoudaki, Oxford University, Oxford, N/A, United Kingdom

1 - When Life Gives You Lemons: Spillovers Between Pharmaceutical Recalls and Shortages

Hanu Tyagi, University of Illinois, Urbana-Champaign, IL, United States, Rachna Shah, Junghee Lee

Operational failures like drug recalls and shortages have serious public health consequences and often occur together. This paper examines how a firm's recall experience affects shortage outcomes - a relationship largely overlooked despite both issues stemming from similar quality and manufacturing problems. Using longitudinal data from 319 firms over 120 months, we find that **completed recalls** reduce the risk of future shortages, while **ongoing recalls** increase the duration of shortages and the risk of future shortages. The effects are stronger when recalls are more severe. We contribute to the literature on learning from failures by showing that **spillovers can occur across different failure types**, not just within the same category. These findings offer actionable insights for managers and regulators on leveraging learning from completed failures while minimizing the harm of ongoing ones.

2 - Strength in Teams: Named Physician Pairs Improve Continuity of Care in Primary Care Amid Workforce Pressures

Laura Poreschack, University of Cologne, Cologne, Germany, Harshita Kajaria-Montag, Stefan Scholtes

Context: Continuity of care - a sustained relationship between a patient and their primary care physician - improves outcomes and reduces healthcare utilization, but workforce pressures threaten its sustainability. Building on the Named Physician model in England, this study proposes a Named Physician Pair model, where two physicians jointly ensure continuity, with a secondary physician covering when the primary physician is unavailable.

Main findings: Using 11 years of data from over 30 million consultations in 381 English primary care practices, we found that seeing a primary physician extends the time to the next consultation by 14.0%, while a secondary physician extends it by 6.5%, compared to other physicians. The benefits of both physicians are particularly strong for older patients, those with multiple comorbidities, and when these physicians are more familiar with the patient and each other. We also observed fewer emergency visits and hospitalizations within a week of consultations with primary or secondary physicians, with secondary physicians achieving more than 50% of the impact seen with primary physicians.

Impact assessment: Applying the Named Physician Pair model to the study practices, targeting the patients who benefit most from continuity, increased consultations with either physician from 53% to 86%, reducing consultation demand by 4.9%. This included 3.4% from primary physicians and 1.5% from secondary physicians, with estimated savings across England exceeding £100 million annually. The study highlights managerial and policy challenges in implementing the Named Physician Pair model but demonstrates its potential to enhance care and alleviate pressures on primary care services.

3 - Short Supply, Long Delays: Evidence from the Parenteral Opioid Shortage

Xinyu Liang, INSEAD, Singapore, Singapore

The U.S. has been grappling with a prolonged opioid epidemic, leading to tighter regulations and production controls; however, as parenteral opioids are critical for acute and surgical pain management, supply disruptions raise serious concerns about unintended consequences for patient care and opioid misuse.

We combine 10+ years of DEA's ARCOS shipment data with patients' electronic health records from Truveta Data. Using geographic variation in shortage exposure, we estimate the impact of a 2017 opioid shortage's impact on care delivery and opioid misuse using a shift-share instrumental variables strategy and an event-study design. We find that the shortage not only delayed care but also increased reliance on riskier opioid use pathways. The findings shed light on how supply shocks in controlled substances affect healthcare delivery and exacerbate risks of opioid misuse.

4 - Improving Rural Healthcare Access: Evidence from Osf Healthcare Facility Expansions in Illinois

Dmitrii Sumkin, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, United States, Ujjal Mukherjee, Sridhar Seshadri

Access to healthcare remains a persistent challenge for rural populations. This study examines how the expansion of OSF HealthCare facilities influences healthcare utilization in rural Illinois, leveraging patient visit data and publicly available social vulnerability indices. We model patient facility choice and implement a staggered difference-in-differences approach, taking advantage of the phased rollout of new healthcare sites across underserved areas. Our analysis finds that the opening of new facilities leads to a significant reduction in the ratio of emergency department visits to regular visits—suggesting a shift toward proactive, preventive care. Additionally, both follow-up visit rates and the average number of follow-up encounters increase, highlighting improved care continuity. These findings demonstrate the vital role of physical access in shaping not only patients' initial care-seeking behavior but also their ongoing engagement with the healthcare system. Based on these insights, we offer actionable, data-driven recommendations to OSF HealthCare for optimizing its network expansion strategy. Specifically, we identify high-impact locations for new clinics, telemedicine hubs, and healthcare kiosks aimed at maximizing benefits for rural communities across Illinois.

SD45

Building B Level 3 B305

Language Models for Public Sector Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Lesley Meng, Yale School of Management, Yale University, New Haven, CT, United States

1 - Reuniting Forcibly Separated Families Through Shared Memories

Huifeng Su, Yale School of Management, New Haven, CT, United States, Lesley Meng, Edieal Pinker

For decades, hundreds of thousands of children have been forcibly separated from their birth families due to abduction or displacement. Many of these separations began in childhood and persisted into adulthood, leaving individuals to search for their families decades later. This situation has created an urgent and large-scale demand for effective reunion services.

These services face significant challenges, including vast search spaces, missing identification information, and incomplete memories. Despite their profound societal importance, family reunification processes remain understudied, particularly with respect to the operational complexities inherent in such searches.

Using data from a large online platform dedicated to family reunification, we develop a predictive pipeline that integrates representation learning with transformer-based language models to recommend likely matches from user-submitted narratives. Our system substantially reduces the search space and, in aggregate, could shorten separation durations by years for affected families. Beyond algorithmic improvements, we also uncover systematic reporting biases in child-submitted reports and demonstrate heterogeneity in how children and parents allocate narrative effort, both of which could undermine the efficiency of searches.

Together, this work advances both the design of algorithmic tools for family reunification and the understanding of human-reporting dynamics that shape their effectiveness.

2 - Keeping the Human in Humanitarian: the Effects of Artificial Intelligence on Donations to Humanitarian Organizations

Marialena Bevilacqua, University of Notre Dame, Notre Dame, IN, United States, Cameron Kormylo, Alfonso Pedraza-Martinez

The increasing adoption of artificial intelligence (AI) by for-profit organizations (FPOs) is largely driven by its potential to enhance operational performance. Nonprofit organizations (NPOs) have begun to adopt AI in hopes of achieving similar benefits. In particular, humanitarian organizations are utilizing AI to improve the efficiency and effectiveness of relief provided to victims of disaster worldwide. However, unlike FPOs that are funded by paying customers, humanitarian organizations rely on donor contributions. Private donors, who do not receive direct goods or services in return for their contributions, are motivated by a range of organizational, internal, and external factors specific to the humanitarian context. While customer preferences are well-studied, little is known about how private donors perceive the use of AI in humanitarian contexts. This research investigates the impact of AI adoption on private donor behavior through a series of online experiments using choice-based conjoint designs. Our findings reveal a strong preference for humanitarian organizations that do not utilize AI for decision making, with significant reductions in donation amounts when AI is involved. This effect holds even when private donors are only exposed to AI-utilizing organizations. Employing a between-subjects experimental design, results suggest an approximately 27% decrease in donations among individuals only exposed to organizations utilizing AI compared to individuals only exposed to organizations not utilizing AI. These insights offer valuable implications for the integration of AI in humanitarian organizations and the importance of addressing private donor perceptions to maintain trust and financial support.

3 - Enabling Personalized Learning with Large Language Models via Reinforcement Learning

Tsai-Hsuan (Angel) Chung, The Wharton School, UPenn, Philadelphia, PA, United States, Hamsa Bastani, Osbert Bastani, Ling-Chieh Kung

Education technology, from Information and Communication Technology (ICT) to recent advancements in Generative AI have proven to be a double-edged sword in shaping student learning, depending on how these tools are incorporated and implemented in education. The inevitable widespread adoption of technologies like ChatGPT highlights the necessity to leverage these tools to enhance rather than hinder educational outcomes. While research indicates that LLMs can democratize education by offering personalized support, the existing personalization approach is “passive”, assuming a level of self-awareness and assessment capability in educators and learners that might not be readily available, especially among low-achieving students and under-resourced settings.

This paper develops and incorporates reinforcement learning with a large language model to help assess each student’s unique educational needs dynamically to provide “auto-adaptive” personalized learning. In collaboration with Department of Education of Taipei City Government, we conduct a large scale randomized control trial (RCT) by offering students a GPT-4-powered mentoring interface enhanced by reinforcement learning. We aim to evaluate whether incorporating reinforcement learning with LLMs significantly enhances the personalization of learning experiences and contributes to more equitable educational environments. Through this study, we demonstrate how adaptive AI interventions can actively support diverse learners and bridge gaps created by traditional passive personalization strategies.

4 - Online, but off-Task: Broadband as a Tool and a Temptation in the Student Learning Process

RJ Niewoehner, Kelley School of Business, Indiana University, Bloomington, IN, United States, Eric Xu, Eunae Yoo, Hyojun Jeong

Background: Providing education is inherently a service operation, an operation with a particular interest in quality outcomes. In today's learning environment, the quality of the service increasingly depends on the quality of the service channel between instructor and student. An internet connection has become the service channel that unlocks resources that can enhance student learning.

Aim: Though prior literature has studied how internet access affects student learning globally, the results remain mixed and context-dependent. We aim to fill the gap by answering: Does access to a reliable internet connection affect student learning outcomes across the US? A longitudinal, US-based study can leverage the diversity between states to connect internet access to student outcomes and so inform national policy and broader international discussions.

Results: All of the following correspond to an increase in math proficiency: greater access to broadband internet, less reliance on cell-only internet, or smaller broadband disparities within a district. Relative to average proficiency, we observe that a 1% increase in broadband corresponds to a greater than 1% increase in proficiency nationally. More interestingly, we find that though proficiency initially improves with broadband speed, the effect eventually reverses (mirroring an inverted U-shape). Such a “tipping point” indicates some students may not be able to resist the temptations of faster internet. We also find some of the effects vary depending on the economic status and even the age of student groupings, as economically advantaged and younger students see greater gains.

5 - Can Increased Platform Liability Reduce Sex Trafficking? Evidence from the Price Effect

Helen Zeng, University of California, Davis, Davis, CA, United States

In March 2018, the U.S. Congress enacted a liability regulation called FOSTA-SESTA that made it illegal for digital platforms to knowingly assist, facilitate, or support sex trafficking. While there has been contentious debate as to whether regulations such as FOSTA-SESTA should exist or be modified to address a variety of concerns, there remains a basic question of whether the law had the intended effect: reducing the supply of minors being advertised on websites providing commercial sex advertisements to US consumers. In this paper, we provide empirical evidence for answering this question.

Our research introduces a novel approach to measuring the effectiveness of platform liability in addressing underage sex trafficking by examining its impact on transaction prices for commercial sex services. Using data from two escort review sites, we employ both linear and double machine learning difference-in-differences frameworks. Our analysis reveals a 7%-17% increase in transacted prices for service providers from the youngest age group relative to other participants. These findings suggest that platform liability increased the difficulty of selling minors in the online commercial sex market. We further show conditions for when this increase in transacted prices of service providers from the youngest age group corresponds to a reduction in the supply of minor participants. Our results contribute to policy discussions on platform regulation through liability-based approaches and offer broader implications for the governance of harmful content in digital marketplaces and the regulation of technological harms.

SD46

Building B Level 3 B306

Strategic Perspectives on Food Supply Chains and Food Security: Allocation, Risk, and Resilience

Invited Session

Decision Analysis Society

Chair: Deniz Besik, University of Richmond, Richmond, VA, United States

Co-Chair: Deniz Karakoc, Arizona State University, Tempe, AZ, United States

1 - A Multi-Tiered Network Equilibrium Model of Fresh Produce Supply Chains with Discounting and Donations

Deniz Besik, University of Richmond, Richmond, VA, United States

This paper develops a multi-tiered network equilibrium model for fresh produce supply chains that incorporates quality deterioration, discount timing decisions, and food donation incentives. Distributors and retailers interact under Cournot-Nash competition, with retailers facing uncertain demand across a finite selling period. Retailers can dynamically adjust prices through discounting and donate surplus produce to food banks to receive tax benefits. The equilibrium model is formulated as a finite-dimensional variational inequality. Numerical results provide insights into optimal pricing, fresh produce allocation, and donation decisions, offering managerial guidance for balancing profit maximization with sustainability and food equity goals.

2 - Natural Disaster Risk and Food Infrastructure

Rocio Tenorio, University of Richmond, Richmond, VA, United States

Food infrastructure is critical for global health and safety. Food banks in particular serve as essential food access resources, as well as for community building, access to health care, and for community disaster preparedness. While these food banks are essential for meeting the needs of vulnerable populations, the food banks themselves can be vulnerable to a variety of natural disaster risks. Following a natural disaster, several factors can influence the food bank’s ability to meet demand, including destruction to the road access network, damage to the facility itself, lack of transportation resources for customers, loss of energy, loss of water, and others. This paper develops a

model to evaluate food infrastructure vulnerability to various disaster risks; and leverages a multi-criteria approach to define highest priorities for investments in food bank resilience. The methods and results of this study will be of interest to risk assessors, policy managers, and the general public.

3 - Spatio-Temporal Trends of Food Consumption: Implications for Food Waste and Insecurity in the U.S

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

Food consumption is a vital topic of discussion regarding the social, economic, and environmental impacts of agri-food systems. Particularly in the United States, food consumption is a major contributor to food waste and a direct indicator of socio-economic inequalities. Thus, it is essential to better understand the country's food consumption trends across time and space in a finer resolution. We introduce a machine-learning integrated multi-criteria decision-making framework to assess how the surrounding environment impacts local food consumption and to what extent. We also pinpoint the spatially resolved highest food-consuming U.S. counties through time. These locations also account for higher food waste and insecurity. Our findings can improve the understanding of agri-food systems and can aid policies to enhance food security and sustainability.

4 - An Optimization Model for Designing Urban Food Networks with Diverse Facility Types and Accessibility Constraints

Meen-Sung Kim, Georgia Institute of Technology, Atlanta, GA, United States, Sofia Perez-Guzman, Carla Tejada, Otgondulam Boldbaatar

Urban food access interventions often rely on limited definitions of infrastructure—focusing primarily on supermarkets—while failing to account for the broader variety of food outlets that serve different community needs. This study develops an optimization framework to guide the design of urban food networks aimed at improving spatial access to fresh food for underserved populations. Building on prior efforts to classify diverse food infrastructure—including traditional and non-traditional outlets with varying purposes—this phase of the project formulates a decision-support model that incorporates accessibility constraints, facility capacity, and minimum consumption thresholds. The model selects from a set of candidate facility locations to ensure each population group is served by at least one viable facility within travel time and distance limits. The objective function incorporates an accessibility measure that balances facility attractiveness—based on capacity—and travel impedance, capturing how distance and travel time reduce the likelihood of access. This allows for exploring trade-offs in placement decisions that affect proximity, coverage, and equitable access. The framework is designed to flexibly incorporate diverse facility types and behavioral access parameters, supporting future exploration of infrastructure configurations in urban settings. The framework is currently being applied to the Chicago metropolitan area to explore how different facility types and spatial configurations affect accessibility outcomes. Its structure supports future integration of diverse food outlet typologies and policy levers, offering a flexible platform for testing equitable strategies to improve urban food access.

5 - A Location-Allocation Model for Food Distribution in Post-Disaster Environments

Diana Ramirez-Rios, University at Buffalo, Buffalo, NY, United States, Angelo Soto-Vergel, Trilce Encarnacion, Johanna Amaya

This research investigates the optimal location decisions in a food distribution network in post-disaster environments. We propose a location-allocation model for a food bank network that minimizes the total social costs generated by distributing food supplies to disaster survivors through the network's points of distribution (PODs), e.g., food pantries. The social costs include the costs of the delivery operation and the external costs of the survivor's suffering in the form of deprivation costs. Our model incorporates an empirically estimated deprivation cost function for food and water supply, which is non-linear with respect to the survivor's deprivation time. We define the deprivation time as the time a survivor has to wait for the delivery, which includes their travel time to the POD and their expected waiting time to receive the food supplies. The model proposes the optimal location of PODs and their allocation to demand zones for food distribution following a disaster, along with a relief distribution strategy. We used the Houston Food Bank network to test the feasibility of activating the nodes as a relief distribution network that could serve the impacted community. The instances aim to mimic a potential design that would have been implemented in response to a future disaster like Hurricane Harvey in 2017. It also addresses the impact of location on delivery frequencies and shipment sizes. The results provide practical insights that can help prioritize the needs of survivors during an emergency or disaster event.

SD47

Building B Level 3 B307

Assorted Topics I

Contributed Session

Contributed

Chair: Justin Keeler, Wichita State University, Maize, KS, United States

1 - A Conjoint Analysis Model with Attribute Groups : Focused on Smartwatch

Sangwon Eum, Korea University, Ansan-si, Korea, Republic of, Hosun Rhim

In this paper, we develop a conjoint analysis model based on attribute groups to effectively consider massive numbers of attributes. In the proposed model, all attributes are divided into multiple attribute groups, and conjoint analysis is performed for each attribute group. Then the results are integrated. Unlike existing models, we do not conduct bridging experiments to aggregate subproblems; instead, we integrate them by including one common attribute to all attribute groups and standardizing based on the part-worths of that attribute. We applied newly developed model to the design of a smartwatch.

2 - After the Party: Unlocking Post Incubation Success in Corporate Incubators

Pragathi Mavinakere Swamygowda, Rotterdam School of Management, Erasmus University, Bengaluru, India, vareska van de rande, Murat Tarakci

Corporate incubators are increasingly adopted by firms to drive innovation by integrating emerging technologies into corporate strategies. However, their true value is realized in the post-incubation phase, where innovations are integrated, scaled, and sustained within the parent firm. Using the structured systematic literature review methodology (Denyer & Tranfield, 2003), we analyzed 40 peer-reviewed articles spanning 2001-2024. Through thematic coding, the study identifies key antecedents, mediators, moderators, and post-incubation innovation outcomes across internal and external corporate incubator models. Internal incubators thrive on R&D capabilities, strategic alignment, and knowledge management systems. These are activated through mediators such as knowledge-sharing platforms and cross-functional collaboration, producing outcomes like accelerated innovation cycles and enhanced market positioning. External incubators, on the other hand, leverage ecosystem connectivity, open innovation practices, and boundary-spanning activities, with mediators including resource fit and external knowledge integration. Their success results in market-oriented innovations and ecosystem synergies but is influenced by moderators such as institutional support and market dynamics.

The study integrates the resource-based view, dynamic capabilities theory, and absorptive capacity theory to contextualize how firms acquire, assimilate and exploit knowledge during post-incubation. Our proposed integrative framework underscores the centrality of absorptive capacity in bridging innovation and application across both incubation models. Thereby, this research provides actionable insights for corporate incubation strategies to ensure sustainable innovation and long-term corporate advantage.

Keywords:

Post-incubation, corporate incubators, innovation management, external incubators, innovation scaling, corporate success, competitive advantage, knowledge integration, strategic agility.

3 - Ecompass: a Decision Support System for Finding Optimal Locations

Vijay Paidi, Dalarna university, Falun, Sweden, Kenneth Carling, Niklas Rudholm

Identifying optimal locations has long been a priority when establishing facilities or business units. Optimal locations can reduce travel distances and, consequently, emissions, contributing to environmental sustainability. While omnichannel retailing and e-tailing are on the rise, a significant number of consumers still visit physical

stores, especially for groceries and other essential goods. This persistence is often due to logistics complexity, automation costs, or the value of in-person product experiences. Optimizing facility locations for consumer electronics, for example, has the potential to reduce emissions by 35%. This opportunity is not unique to retail, it extends to sectors such as healthcare, education, and urban planning. Given the widespread relevance of this issue, there is a clear need for user-friendly, open-source tools that support such location decisions. We present eCompass, an open-source web application that helps identify optimal facility locations by considering both population density and associated carbon emissions. eCompass can support decisions such as selecting a new store location, evaluating a set of candidate sites, expanding a business chain, or pruning underperforming outlets. We demonstrate its utility using a case study of IKEA store locations in Sweden. In 2013, IKEA operated 17 stores and expanded to 20 by 2015. Our analysis shows that similar access could have been achieved with fewer stores, highlighting how eCompass helps to reduce emissions by minimizing unnecessary travel. Although eCompass does not address all real-world complexities of facility location, it offers valuable, data-driven insights to support decision-making in sustainability-conscious planning.

4 - Geodesic Paths and Location Problems in Different Lp-Norms Polyhedral Complexes

Justo Puerto, Universidad de Sevilla, Seville, Spain

In this paper we address two different related problems. We first study the problem of finding a simple shortest path in a d-dimensional real space subdivided in several polyhedra endowed with different lp-norms. The second problem that we consider is the continuous single facility location or Weber problem that results in this subdivision of lp-normed polyhedra. The first problem is a variant of the weighted region problem, a classical path problem in computational geometry introduced. As done in the literature for other geodesic path problems, we relate its local optimality condition with Snell's law and provide an extension of this law in our framework space. We propose a solution scheme based on the representation of the problem as a mixed-integer second order cone problem (MISOCp). We derive two different MISOCp formulations, theoretically compare the lower bounds provided by their continuous relaxations, and propose a preprocessing scheme to improve their performance. The usefulness of this approach is validated through computational experiments. The formulations provided are flexible since some extensions of the problem can be handled by transforming the input data in a simple way. To solve the second problem, we adapt the solution scheme that we developed for the shortest path problem and validate our methodology with extensive computational experiments.

5 - Sorting Matters: On Exact Recovery in Discrete Ordered Median Problems

Victor Blanco, University of Granada, Granada, Spain, Diego Laborda, Miguel Martinez-Anton

The core goal of clustering is data group recognition according to some criterion. It ensures a non-supervised classification of data into clusters with similar features. Because of its practical potential, clustering problems have become a main concern in data science and a ubiquitous tool with a wide range of applications in natural or social science, management, and engineering. Among the goodness of this class of problems, one finds the flexibility and relevance of the decision-maker at the time of setting the group criterion. Over the last decades, modern optimization has emerged as a powerful field to tackle problems in classification and data science. Within this framework, clustering problems meet location theory. The k-median clustering approach uses a proximity criterion to detect data groups close to each other and distant from others.

In 2000, Puerto and Fernandez introduced ordered median operators into location theory, which serve as clustering criteria and give rise to the class of discrete ordered median problems (DOMP). These aim to minimize the weighted sum of distances to their respective centers, accounting for the order of distances, making sorting part of the decision. By adjusting the weights, one can retrieve the k-median, the minimax, among other robust variants.

We study exact recovery in key mixed-integer linear programs for DOMP, identifying conditions that guarantee optimal integer solutions for their linear programming (LP) relaxations. This means a first step toward exact and polynomial-time solutions via LP relaxation, enabled by prior knowledge of data distribution in ordered problems.

6 - Don't Smash the Watermelon! Field-Tested Tools for Teaching Analytics Field Courses

Patrick Noonan, Emory Univ., Emeritus, Greensboro, NC, United States

Project-based field courses - real problems with real clients - are widely used to bridge theory and practice, to infuse soft skills into a hard curriculum, and to otherwise integrate and reinforce diverse methodological approaches. But these courses are hard to pull off well, with risks for students, clients and the faculty responsible for them. A field-tested curriculum built around "structured problem solving" (the canon of the consulting profession), along with a classroom-tested pedagogy for analytics students, can increase the effectiveness of field projects and better prepare students for the work environment.

7 - Beautifying Analytical Workflow and Syntax Documentation via Quarto: a Technical Workshop Session

Justin Keeler, Wichita State University, Maize, KS, United States

This technical session is designed to show scholars how to take their workflow and analytical syntax documentation beyond scripts in an IDE. A focus will be on using Quarto within RStudio for R and within VSCode for Python. I will provide access to my GitHub repository for attendees to have working code as templates during and after the session. I need a full 75 minutes for this workshop. The value proposition of this offering is showing individuals how to create html and/or pdfs of their workflows/syntax can have multiple purposes - personal annotation, teaching documentation, or potential book publishing (see R for Data Science - <https://r4ds.hadley.nz/quarto.html>).

SD48

Building B Level 3 B308

Model Building for Better Decision Making

Invited Session

Decision Analysis Society

Chair: Andrea Hupman, University of Missouri - St. Louis, Saint Louis, MO, United States

1 - *Nr Qflex: a Quantile Basis Expansion Framework for Flexible Probability Distributions

Eric Bickel, The University of Texas at Austin, Austin, TX, United States, Connor Colombe, Benjamin Leibowicz

We present **QFlex**, a new family of probability distributions built on the **Quantile Basis Expansion** (QuBE) framework. QuBE defines quantile functions as linear combinations of monotonic basis functions composed with valid base quantile functions, ensuring strict monotonicity and distributional validity by construction. QFlex specializes this approach using interleaved odd-power transformations of exponential, reflected exponential, and uniform quantile functions. The resulting system offers high flexibility, supports bounded and semi-bounded forms via transformations, and allows interpolation through arbitrary quantile assessments. Compared to existing systems like Metalog, QFlex demonstrates improved accuracy, stability, and structural interpretability across a wide range of distribution shapes and applications.

2 - Developing a Value Model Using Microsoft Copilot & Potential and Pitfalls

Gregory Parnell, University of Arkansas, Fayetteville, AR, United States, Eric Specking, Nicholas Shallcross

Unmanned Aeronautical Vehicle (UAV) Swarms are being used and developed for many national security and civilian applications. We decided to develop a multiple objective value model for two book chapters in a new book on Military Operational Analysis and use UAV Swarms as our illustrative example. One chapter would develop the model. The second chapter would apply the model to an analysis of alternatives. As an experiment, we decided to use Microsoft CoPilot prompts as the primary source for the objectives, value measures, and value model. We describe the prompts we used, the CoPilot responses, and the changes the authors made to correct omissions and errors. The results demonstrate the potential and the pitfalls of using generative AI in decision analysis

3 - Wisdom of Sequential Crowds: Quantile Forecast Aggregation with Learning and Updating

Majid Karimi, California State University San Marcos, San Marcos, CA, United States, Arthur Carvalho

From forecasting the yield of corn in agribusiness problems to winning the Nenana Ice Classic by correctly estimating the exact date and time the ice breaks on the Tanana River, point forecasts are often made sequentially, allowing for *learning* and *updating*. This study investigates quantile forecast aggregation in the presence of learning from previous forecasts and updating whenever new information becomes available. In the absence of learning or updating, simple averaging of quantiles is often a common and robust solution, canceling out the individual biases. In sequential settings, however, point forecasters (humans, machines, or both) can engage in social learning — learning from others' actions — and update their forecasts in light of new information. Furthermore, optimal forecast combinations in sequential settings, in particular, can be hard to come by due to the significance, relevance, or efficacy of the forecasters' historical performance. Using synthetic and real data, we explore different aggregation methods, including our proposed Bayesian approach, and compare different techniques. We find that incorporating learning and updating results in sharper and more calibrated forecasts in sequential settings, particularly for the proposed Bayesian approach and even when considering simple learning heuristics such as quantile matching estimation.

4 - Combining Model Predictions for Decisions with Asymmetric Costs

Andrea Hupman, University of Missouri - St. Louis, Saint Louis, MO, United States

Predictive models provide information to help determine policies that maximize the value to the decision maker. This work examines the case of a two-action decision that is informed by an ensemble of predictions from multiple models, and where the prediction errors have asymmetric costs. We examine how the decision maker's risk attitude and the underlying probability distributions interact to affect value, and also how the properties of the component models and the approach to combining them interact to affect value.

SD49

Building B Level 3 B309

Experimentation in Operations

Invited Session

MSOM: Supply Chain

Chair: Chonghuan Wang, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Effect Sizes, Sequential Methods, and Advancing Experimentation for Science and Industry

Ken Kelley, University of Notre Dame, Notre Dame, IN, United States

Experimentation in digital and non-digital contexts is about making inferences from samples to a population. Inferences may be in terms of the existence of effects, the size of effects, finding the best treatment, among other reasons. In this talk I discuss questions a manager or research might have when designing studies and discuss multiple perspectives of inferences for multiple types of effects. In particular, I consider planning studies from the accuracy in parameter perspective to obtain accurate estimates of magnitude of an effect or statistical power analysis to have a specified probability of obtaining a statistically significant effect if the null hypothesis is false. With regards to effect sizes, in addition to discussing the concept of effect size itself, I consider omnibus effects to quantify a general hypothesis and targeted effect sizes to quantify a specific effect of interest. I show how different types of designs for digital experimentation have implications that go beyond a single experiment and discuss how a narrow view of experimentation can impact the way in which a firm leverages study results. Ultimately, I argue that while selection of the best of two treatments (A/B) might be an industry standard, it is a narrow way to understand customers, contexts, and limits a firm's cumulative knowledge and consideration of heterogeneous findings that can lead to future advancement and successes.

2 - Safe Reinforcement Learning with Contextual Information: Theory and Applications

Esmail Keyvanshokoo, Mays Business School Texas A&M University, College Station, TX, United States, Junyu Cao, Tian Liu

We develop a new safe reinforcement learning (RL) theory that learns a personalized, safe control policy under uncertainty, modeled as a contextual constrained Markov decision process with unknown dynamics and constraints. Our approach ensures safety throughout the learning process and achieves sub-linear regret, significantly improving upon existing results in both safe and contextual RL. We validate our approach using synthetic and clinical data from patients with co-morbid type 2 diabetes and hypertension, demonstrating its superiority over current benchmarks and highlighting its broad applicability to domains requiring both safety and personalization.

3 - Pricing Experiments in Matching Marketplaces Under Interference: Designs and Estimators

Arthur Delarue, UVA Darden School of Business, Charlottesville, VA, United States, Kleanthis Karakolios

Interference between treated and untreated units is a source of bias in marketplace experiments. In this paper, we specifically consider pricing interventions, in which a platform seeks to adjust base pricing levels at the marketplace level in order to increase demand. In a matching marketplace, this type of experiment leads to a crucial design question: should the platform match treated and untreated units differently because they paid different prices? We find that standard estimation techniques are biased, but the sign of this bias depends strongly on this design choice. Bias can be reduced by using the "shadow price estimator", which relies on the optimal dual solution of the platform's supply-demand matching problem; especially when the platform chooses to ignore pricing differences at matching time. We validate our findings both theoretically in a fluid limit setting, and numerically in a finite-sample setting.

4 - Benefits and Costs of Adaptive Sampling for Inference

Yu-Shiou Lin, Columbia University, New York City, NY, United States, Dae Woong Ham, Iavor Bojinov

Multi-arm bandits are gaining popularity as they enable real-world sequential decision-making across application areas, including clinical trials, recommender systems, and online decision-making. While prior work typically focus on regret-minimizing properties of multi-arm bandits or enabling valid inference from adaptively collect data, it is still unclear whether one should adaptively collect data for more effective inference. To address this gap, we characterize the trade-offs involved in adaptive experimentation when both objectives are considered in a multi-arm setting.

Using both fixed and adaptive allocation rules, we analyze different mean squared error (MSE) based objectives as a measure for inference guarantees. Our theoretical results demonstrate that adaptive allocation can reduce total MSE when variances are heterogeneous and correlated with arm means; however, no single allocation rule uniformly dominates across all arms. This finding underscores the importance of optimizing inference for a targeted subset of arms rather than uniformly distributing samples across all treatments. Simulation results support these insights: adaptive strategies consistently outperform uniform allocation for high-variance arms, particularly in settings with imbalanced variance. These findings motivate a reevaluation of the goals of online experiments. Beyond traditional regret minimization, we advocate for inference-oriented performance metrics that also consider estimation accuracy for arms of greatest interest. This work proposes a framework for adaptive experimentation that integrates exploitation, statistical efficiency, and targeted inferential precision.

SD50

Building B Level 3 B310

Socially Aware Service Operations

Invited Session

MSOM: Service Operations

Chair: Vahideh Manshadi, Yale, Westwood, United States

1 - Who to Offer, and When: Redesigning Feeding America's Real-time Donation System Amid Heterogeneous Response Times

Soonbong Lee, Yale university, New Haven, CT, United States, Vahideh Manshadi, Daniela Saban

In collaboration with Feeding America, we aim to redesign RealTime, its growing food rescue system that connects ad-hoc, time-sensitive food donations to local agencies through an offering process. Due to heterogeneity in agencies' response rates, in making offering decisions, RealTime and similar platforms face a key challenge in balancing efficiency and equity: offering to many agencies upfront improves efficiency (likelihood of acceptance) but may disadvantage those with lower response rates. The uncertainty of future donations further complicates offering decisions. Motivated by these challenges, we study a dynamic offering problem in which donations arrive sequentially and are matched through a multi-stage offering process. The goal is to maximize an objective balancing efficiency and equity—promoting allocations proportional to agencies' needs. We first develop a dynamic programming (DP) algorithm that optimally solves the one-donation problem and yields an intuitive nested offering sequence. Moving beyond one donation, and motivated by the ad-hoc nature of donations, we take a robust approach to designing sequential offering policies that do not rely on knowledge of future arrivals. Our policy solves a modified one-donation DP by applying properly designed penalties to historical allocations to hedge against uncertainty. We establish strong—and optimal in asymptotic regimes—performance guarantees of our proposed policy. We further show the importance of hedging: a greedy policy solving one-donation DP without hedging fails to achieve comparable guarantees. Numerical results on real data demonstrate that our approach significantly improves both equity and efficiency relative to current practice and several benchmarks.

2 - Algorithmic Pricing, Transparency, and Discrimination in the Gig Economy

Daniel Chen, Boston College, Chestnut Hill, MA, United States, Gad Allon, Kenneth Moon

Algorithms control pricing and match customers and workers in the gig economy. However, algorithms face several critiques: they lack transparency, can be biased, and can be inefficient. We empirically analyze these issues and show that variation in algorithmic pricing incentivizes workers to decline a large portion of trips in practice, significantly decreasing system efficiency. We estimate policy counterfactuals and show that minimum wage and transparent pricing policies can improve efficiency without harming fairness constraints.

3 - Data-Driven Inventory Management Under Privacy Concerns

Elena Belavina, Cornell University, New York, NY, United States, Lorraine Yuan

Firms increasingly rely on customer data to improve demand forecasting and inventory management. However, attackers can exploit these data-driven decisions—by linking them with external information—to infer sensitive data, raising serious privacy concerns and limiting adoption. This study develops and compares privacy-preserving approaches for data-driven inventory management that are robust to such attacks. We propose differential privacy-based adaptations of two leading data-driven inventory management methods—the traditional two-step estimate-then-optimize, and a more contemporary one-step joint estimate-optimize paradigm—and establish theoretical guarantees on their performance. We quantify trade-offs between privacy protection, profitability, and customer surplus. Our analysis shows that the one-step paradigm dominates the two-step paradigm by enabling lower-cost privatization of customer data through more targeted noise injection, achieving lower privacy exposure risk, higher profitability, and greater customer surplus. This finding reshapes our understanding of privacy protection—the costs and benefits of privacy preservation are not only driven by the privacy algorithms used but are also fundamentally shaped by the underlying data-driven decision-making paradigm. We also show that a firm that uses our methods and chooses an optimal privacy level will, in fact, increase its demand, profit, and customer surplus (under most conditions)—challenging the notion that privacy protection comes necessarily at a cost. Numerically, one-step algorithms offer 3–6 stronger privacy and up to 40% higher profits than the two-step alternative—benefits that remain even under high customer sensitivity to privacy.

4 - Multi-Listing for Horizontally Differentiated Services

Yichuan Ding, McGill University, Montreal, QC, Canada, Zhou Chen, Luyi Yang

This paper studies a queueing system with horizontally differentiated servers. Each customer decides whether to join the queues of multiple servers simultaneously ("multi-list"), join the queue exclusively served by her preferred server, or balk. Such a multi-listing system contrasts a single-listing system, which allows for server choice but prohibits multi-listing, and a pooling system, which precludes server choice by requiring every joining customer to multi-list. We build queueing-game-theoretic models of the three systems above and compare their throughput and social welfare in equilibrium. Between a single-listing system and a pooling system, the former better facilitates matching between customers and servers, whereas the latter excels at reducing waiting time through load balancing. Accordingly, we find that either system can outperform the other in either throughput or social welfare. We find that multi-listing indeed outperforms pooling in both throughput and social welfare. While multi-listing also achieves higher throughput than single-listing, it can strikingly underperform single-listing in social welfare. In contrast to the equilibrium, both the social-welfare-maximizing or revenue-maximizing routing policies of the multi-listing system can be asymmetric across the servers even when the servers are ex-ante symmetric. Relative to the socially or revenue-optimal state, customers under-choose multi-listing in equilibrium when the congestion level is low but may over-multi-list otherwise. The social planner can charge nonnegative, asymmetric prices to restore efficiency and profitability in equilibrium. Our paper provides design guidance for the configuration of multi-server service systems.

SD51

Building B Level 3 B311

Data-driven Sustainable Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Neha Sharma, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States

1 - Modeling Borrower Intent: Optimizing Commercial EV Loan Management

Leann Thayaparan, Johns Hopkins University, Baltimore, MD, United States, Vikas Deep, Neha Sharma

The adoption of commercial electric vehicles (CEVs) is crucial for reducing logistics emissions but remains limited, especially among small businesses that dominate markets like India. These firms face high upfront costs and lack access to favorable financing due to uncertainties around CEV value and borrowers' creditworthiness.

This study presents a novel framework to improve lending decisions and interventions by introducing a latent "willingness to pay" (WTP) metric. We model borrower behavior using a Markov decision process, where the firm's actions—calls, visits, or repossession—depend on WTP estimates derived from high-dimensional features including vehicle usage.

Our collaboration with Turno, a major Indian CEV lender, provides a rich dataset combining financial and usage information. Results show our approach improves prediction accuracy, reduces bias, and better allocates costly interventions. By explicitly modeling WTP, we also introduce interpretable and fairer decision-making that considers not just repayment likelihood but also the cost-benefit tradeoffs of lender actions.

This work contributes to literature on credit risk, fairness, and interpretable machine learning. It highlights the value of data sharing between EV producers and lenders, offering a practical path to broaden CEV adoption while maintaining financial sustainability in emerging markets.

2 - Learning Sustainability Assessments

Ilgın Dogan, INSEAD, Fontainebleau, France, Atalay Atası, Luk Van Wassenhove

This paper studies the ecosystem of a sustainability rating company as a sequential triadic game involving buyers, suppliers, and the company. Each party engages in parallel learning: buyers and suppliers assess the ratings' accuracy in capturing long-term risks which in turn influences their retention, while the company learns their behaviors to adapt pricing and maximize system-wide retention. Calibrated with real data, our model provides data-driven, regret-minimizing sequential decision policies that help all three parties reduce ambiguity in measuring sustainability risks.

3 - Optimal Experimental Design with Routing Constraints

Baizhi Song, London Business School, London, United Kingdom, Jean Pauphilet

Data collection in application domains like agriculture and environmental science requires the deployment of sensors in large remote areas. These use cases challenge the traditional optimal experimental design (OED) formulation from statistics by their scale as well as the presence of complex operational constraints, such as that data is collected along the trajectory of a drone. This paper proposes a tailored branch-and-bound algorithm to solve large-scale instances of OED as well as account for routing constraints. Efficient screening rules and first-order methods for solving the relaxations are key to the tractability of our algorithm, which significantly outperforms commercial solvers on both the classical OED problem and OED with routing constraints. For classical OED for example, it increases the size of instances solvable under a minute from $n=200$ to $n=10,000$. In the presence of routing constraints, it finds solutions with up to 10% higher objective value and solves twice as many instances to optimality. In use cases where data is collected to construct a predictive model, our approach achieves the same accuracy as alternative collection strategies with 30-50% less samples, hence demonstrating the potential impact of optimization on data collection cost.

4 - A Dual Perspective on Decision-Focused Learning

Paula Rodriguez Diaz, Harvard University, Boston, MA, United States, Kirk Bansak, Elisabeth Paulson

Many decision-making pipelines solve an optimization model on predicted inputs (predict-then-optimize, PtO). Decision-Focused Learning (DFL) trains predictors to improve downstream decisions rather than minimize predictive error; however, state-of-the-art methods either differentiate through the optimizer or employ tight task-specific surrogates, requiring frequent solver calls and inflating training cost. We propose a dual-guided DFL procedure that preserves decision alignment while sharply reducing solver-in-the-loop cost: the downstream problem is solved only at scheduled refreshes to estimate Lagrange multipliers, and between refreshes the predictor is updated on dual-adjusted targets using simple surrogate losses. Tuning the refresh interval trades compute for fidelity: longer intervals approach standard ERM, while shorter ones recover classical DFL. Across representative PtO tasks, the method matches the decision quality of solver-heavy baselines with substantially fewer solver calls and lower wall-clock time, improving the scalability of decision-aware learning to large datasets.

SD53

Building B Level 3 B313

Consumer Choice Models

Invited Session

Revenue Management

Chair: Qi Feng, Purdue University, West Lafayette, IN, United States

Co-Chair: MENGYING XUE, Fudan University, Shanghai, N/A

1 - A Feature-Based Consideration Set Choice Model for Online Retailing

Ashwin Venkataraman, University of Texas at Dallas, Richardson, TX, United States, Amin Farzaneh, Sajad Modaresi

We propose a feature-based consideration set choice model (FCM) motivated by customers' purchasing behavior in online retail platforms. The FCM incorporates heterogeneity in customer preferences across both the consideration and purchase stages, and allows customers to re-evaluate products' utility in the second stage. We develop an efficient maximum likelihood estimation procedure for estimating the model parameters from customers' click and purchase data. We show that the assortment optimization problem under FCM is NP-hard by drawing a connection to the assortment problem under the LC-MNL model. We propose a variant of our model that admits a tractable MILP formulation for solving the assortment problem and also facilitates the study of the more complex joint assortment and pricing problem. Through numerical experiments on real and synthetic data, we demonstrate that our proposed model outperforms the MNL and LC-MNL benchmarks on out-of-sample prediction accuracy and decision performance. Finally, we introduce a novel operational problem which identifies the subset of features to display during the consideration formation stage to maximize expected revenue. We establish the NP-hardness of this problem under both model variants and propose a tractable MILP formulation for solving it.

2 - Modeling Consumer's Rationality: the Token Tree

Qi Feng, Purdue University, West Lafayette, IN, United States, J Shanthikumar, MENGYING XUE

We propose a tree-based representation to model individual consumer choices over a set of available options by allocating tokens along the tree branching based on the choice outcomes. The token tree framework is flexible enough to capture both rational and boundedly irrational behaviors in consumer decision-making. At the aggregate level, the expected token tree provides a basis for empirically testing the rationality of consumer choices and identifying meaningful clusters of consumer segments.

3 - Approximation Schemes for Dynamic Pricing with Opaque Products

Xingxing Chen, University of Richmond, Richmond, VA, United States, Jacob Feldman, Danny Segev, Levi DeValve, Yukai Huang

In this paper, we consider a multi-period, multi-product dynamic pricing problem in which each product is endowed with an exogenous starting inventory level, and there is the added complexity of an opaque selling option. That is, alongside traditional (transparent) products, the retailer or platform also has the option to create and price an opaque product, which corresponds to a dummy product comprised potentially of any subset of the displayed transparent products. In the event that a customer selects the opaque product, the platform has the freedom to choose any of the opaque product's constituents to satisfy this demand. All-in-all, we are left with a classical dynamic pricing problem with a twist, since the addition of the opaque selling option gives the platform an extra lever of flexibility to balance supply and demand. For this novel product setting, we develop a nuanced constant factor approximation scheme for choosing the opaque product as well as the dynamic transparent and opaque product prices in a multi-period setting.

SD54

Building B Level 2 B216

Topics in Emerging Technologies and Services

Invited Session

Revenue Management

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

1 - Going Live, Going Unequal? Gender Gaps in Patient Demand on Online Medical Platforms

YIXUAN LIU, Hong Kong Polytechnic University, Hong Kong, Hong Kong, YIXIN LU, Xiaofei Zhang

Medical live streaming has emerged as a novel communication tool on online medical consultation (OMC) platforms, enabling physicians to engage patients through real-time video broadcasts. However, its impact on physician inequality—particularly gender-based disparities in patient demand—remains understudied. In particular, while medical live streaming may enhance patient engagement and visibility, concerns persist that it could amplify pre-existing gender biases by disproportionately favoring male physicians, thereby exacerbating inequality in how patients allocate attention. This raises important questions about whether digital technologies democratize access or reinforce structural disparities on OMC platforms. Using data from a leading Chinese OMC platform, we examine how medical live streaming affects the gender gap in patient demand. We find that live streaming significantly increases patient demand for both male and female physicians, with a larger gain for male physicians—thereby widening the platform's existing gender gap. However, this effect reverses in female-dominated specialties and is moderated by physician seniority and streaming characteristics. Female physicians benefit more when holding junior seniority or streaming in environments that convey clinical professionalism, while male physicians gain more when exhibiting stronger emotional variation. Our study highlights how emerging digital tools can either alleviate or amplify structural inequalities depending on their design and usage. We offer theoretical and practical implications for platform operators seeking to promote equitable patient access while enhancing overall engagement through medical live streaming.

2 - Fintech Lending, Open Banking, and Data Manipulation

Yangguang Wang, University of Illinois Urbana-Champaign, Champaign, IL, United States, Fasheng Xu, Zhengping Wu

Financial inclusion remains a major challenge due to information asymmetry and limited credit histories, especially for underserved borrowers. While FinTech firms have broadened credit access using alternative data, such data can be strategically manipulated by consumers. This raises concerns about whether open banking (OB)—which grants access to consumers' banking data—can mitigate these issues and create value. We develop a signaling model to analyze OB's impact on consumer manipulation, financial inclusion, and market efficiency. OB introduces two opposing forces: a *muddling effect* that undermines alternative data value, and a *differentiating effect* that improves screening accuracy. The dominance of either effect depends on manipulation costs. When the muddling effect prevails, OB can reduce FinTech profitability and hinder inclusion. Still, it consistently raises consumer surplus by balancing manipulation incentives and lending decisions. To enhance OB's effectiveness, we propose two interventions: (1) **calibrated algorithmic transparency** to steer manipulation in beneficial ways, and (2) **semi-open banking**, which selectively limits banking data access to improve screening. These strategies help FinTechs restore data value, increase profitability, and promote broader financial inclusion. Our findings inform FinTech and policy efforts to design OB initiatives that better align consumer behavior with market efficiency.

3 - Swap Contract for Production Motivation in General Risk-Averse Newsvendor Problems

Yiwen Shen, HKUST, Hong Kong, Hong Kong, Qingfa Zhang

Risk-aversion plays a central role in agent's decision-making. In particular, it can drive producers to set lower production levels, causing economic inefficiency and welfare loss. We study this problem in a general newsvendor setting, with a social planner aiming to incentivize production from risk-averse producers. We model the decision-making of producers by a general optimization problem, which encompasses most of the objectives and risk constraints considered in the literature. To mitigate underproduction, we develop a novel swap contract between the social planner and producer by extending a return payoff function proposed in recent operations management literature. For producers with general optimization problems, the swap contract increases their optimal production quantities in an incentive compatible way towards the risk-neutral level. The swap contract is budget neutral for the social planner and robust to information asymmetry regarding producer's concrete optimization problem. Economically, the swap contract decouples the operational decision from the risk preference of producers. In addition, the social planner can manage his risk associated with the swap contract by pooling and trading his claims.

4 - Persuasion and AI News Summary

Jiding Zhang, Arizona State University, Tempe, AZ, United States, Lin Hu, Katsiaryna Siamionava, Shuang Gao, Pei-yu Chen
TBD.

SD55

Building B Level 3 B314

Societal Aspects of Markets -- Vahideh & Faidra

Invited Session

Revenue Management

Chair: Vahideh Manshadi, Yale, Westwood, United States

Co-Chair: Faidra Monachou, Yale University, New Haven, CT, United States

1 - Why the Rooney Rule Fumbles: Limitations of Interview-Stage Diversity Interventions in Labor Markets

Faidra Monachou, Yale University, New Haven, CT, United States, Setareh Farajollahzadeh, Soonbong Lee, Vahideh Manshadi

Many industries, including NFL with Rooney Rule and law firms with Mansfield Rule, have adopted interview-stage diversity interventions of inconclusive effectiveness. Motivated by this, we develop a framework of a two-stage hiring process, where rational firms, with limited interview and hiring capacities, aim to maximize the match value of their hires. There are two equi-sized groups, m and w , with identical ex-post match value distributions. Match values are revealed only post-interview, while interview decisions rely on partially informative pre-interview scores. Pre-interview scores are more informative for group m , while interviews reveal more for w . Therefore, if firms could interview all candidates, both groups would be equally hired. However, we show that requiring equal representation in the interview stage does not translate into equal representation in hiring. With or without intervention, a firm may interview more group w candidates but still hire fewer. At an individual level, we show that strong candidates from both groups benefit from the intervention as the candidate-level competition weakens. For borderline candidates, only group w candidates gain at the expense of group m . Furthermore, we study two vertically differentiated firms, where only the top firm adopts the intervention. We show that, due to increased firm-level competition, the lower firm may hire fewer group w candidates, and find examples where fewer group w candidates are hired across the market. At an individual level, while superstar candidates in both groups benefit, surprisingly borderline candidates may be worse. Generally, our natural two-stage market framework may be of independent interest.

2 - Centralization Vs Decentralization in Hiring and Admissions

Juba Ziani, Georgia Tech, Atlanta, GA, United States, Benjamin Fish, Diptangshu Sen

There is a range of ways to organize hiring and admissions in higher education, as in many domains, ranging from very centralized processes where a single person makes final decisions to very decentralized processes where many people make decisions about who to admit or hire. Decentralized processes can enable individual and collective empowerment, but this may come at the cost of efficiency. With the advent of automated decision making, this question of centralization has a big impact on hiring and admissions, given that automated systems often are easier to implement, or even require, more centralized decision making.

In this paper, we develop a strategic model to explore the impact of the degree of centralization on both the candidates and the hirers, with a focus on university admissions. The model reflects a trade-off between a centralized committee where preferences may not capture individual hirers' preferences, and a decentralized process where individual hirers face extra costs to interview candidates themselves. We characterize when individual hirers prefer the decentralized process over the centralized process as a function of the degree to which the centralized process and hirers' preferences are aligned. We also show that decentralization can have devastating consequences for fairness, leading to major disparities in the likelihood of getting hired across candidates. Our results demonstrate the trade-offs that occur under the question of centralization vs decentralization, and point to how an answer to this question can impose significant harm to people in these systems.

3 - Learning Fair Demand Models

Hyemi Kim, Columbia University, New York, NY, United States, Adam Elmachtoub, Jonathan Tan

Fairness is an increasingly critical concern in algorithmic pricing, particularly when decisions have disparate impacts across demographic groups. This work examines the role of fairness constraints in machine learning-based pricing pipelines, focusing on a two-stage framework comprising demand estimation and price optimization. We incorporate fairness into the estimation stage from both price and demand perspectives and assess its impact on downstream outcomes such as revenue. These results are then compared to interventions applied at the price optimization stage, highlighting how the point of fairness intervention can influence downstream outcomes.

4 - Adaptive Priority Mechanisms

Oguzhan Celebi, University of Michigan, Ann Arbor, MI, United States, Joel Flynn

How should authorities that care about match quality and diversity allocate resources in the face of uncertainty? We introduce adaptive priority mechanisms (APM) that prioritize agents based on their scores and characteristics. We show that APM uniquely implement the ex post optimal allocation. The ubiquitous priority and quota mechanisms are optimal if and only if the authority is risk-neutral or extremely risk-averse over diversity, respectively. Deferred Acceptance implements the unique stable

matching when all authorities use the optimal APM. We provide a practical roadmap for implementing APM as a market-design solution and illustrate this using Chicago Public Schools data.

5 - Accelerated Allocation Policies to Improve Utilization of Marginal Deceased Donor Kidneys

Grace Guan, Stanford University, Stanford, CA, United States, Felipe Subiabre, Philipp Strack, Itai Ashlagi

While around 90,000 patients are currently waiting for a deceased donor kidney transplant, almost 30% of recovered deceased donor kidneys go unused due to inefficiencies in the current allocation system. The Organ Procurement and Transplantation Network is currently searching for a structured and equitable way to improve organ utilization. We demonstrate how data-driven insights and simulation can provide a transparent and principled way to identify which organs are at risk of nonuse and how to allocate them.

SD56

Building B Level 3 B315

Auctions under Limited Information

Invited Session

Auctions and Market Design

Chair: Rachitesh Kumar, Carnegie Mellon University, Pittsburgh, PA, United States

Co-Chair: Omar Mouchtaki, New York University, New York, NY, United States

1 - Single-Sample Online Resource Allocation

Sahil Singla, Georgia Tech, Atlanta, GA, United States

In online resource allocation, n buyers arrive sequentially, each with a combinatorial valuation—such as submodular or XOS—over subsets of m indivisible items. The goal is to immediately allocate a subset of the remaining items to maximize total welfare, defined as the sum of buyer valuations. While extensive prior work has addressed this problem under the assumption that buyer valuations are drawn from known, independent distributions, these algorithms typically rely on full access to the underlying distributions.

In contrast, real-world applications often offer only limited information. In this talk, we investigate how well one can approximate the hindsight optimum when the algorithm is given just a **single sample** from each buyer's valuation distribution. We show that even this minimal input suffices to achieve a **(1- ϵ)-approximation** for general buyers, assuming large item budgets, and an **$\Omega(1)$ -approximation** for submodular and XOS buyers even with small budgets.

This talk is based on joint work appearing in **STOC 2025** and **FOCS 2024**.

2 - Auction Design Using Value Prediction with Hallucinations

Omar Mouchtaki, New York University, New York, NY, United States, Ilan Lobel, Humberto Moreira

We investigate a Bayesian mechanism design problem where a seller seeks to maximize revenue by selling an indivisible good to one of n buyers, incorporating potentially unreliable predictions (signals) of buyers' private values derived from a machine learning model. We propose a framework where these signals are sometimes reflective of buyers' true valuations but other times are hallucinations, which are uncorrelated with the buyers' true valuations. Our main contribution is a characterization of the optimal auction under this framework. Our characterization establishes a near-decomposition of how to treat types above and below the signal. For the one buyer case, the seller's optimal strategy is to post one of three fairly intuitive prices depending on the signal, which we call the "ignore", "follow" and "cap" actions.

3 - An Equilibrium Model of the First-Price Auction with Strategic Uncertainty: Theory and Empirics

Bernhard Kasberger, University of Linz, Linz, Austria

In many first-price auctions, bidders face considerable strategic uncertainty: they cannot perfectly anticipate the other bidders' bidding behavior. We propose a model in which bidders do not know the entire distribution of opponent bids but only the expected winning bid and lower and upper bounds on the opponents' bids. We characterize the optimal bidding strategies and prove the existence of a novel equilibrium. Finally, we apply the model to estimate the cost distribution in highway procurement auctions. Our model fits the data better than existing approaches in which the entire distribution of opponent bidders is assumed to be known.

4 - Prior-Independent Bidding Strategies for First-Price Auctions

Rachitesh Kumar, Carnegie Mellon University, Pittsburgh, PA, United States, Omar Mouchtaki

First-price auctions are one of the most popular mechanisms for selling goods and services, with applications ranging from display advertising to timber sales. Unlike their close cousin, the second-price auction, first-price auctions do not admit a dominant strategy. Instead, each buyer must design a bidding strategy that maps values to bids -- a task that is often challenging due to the lack of prior knowledge about competing bids. To address this challenge, we conduct a principled analysis of prior-independent bidding strategies for first-price auctions using worst-case regret as the performance measure. First, we develop a technique to evaluate the worst-case regret for (almost) any given value distribution and bidding strategy, reducing the complex task of ascertaining the worst-case competing-bid distribution to a simple line search. Next, building on our evaluation technique, we minimize worst-case regret and characterize a minimax-optimal bidding strategy for every value distribution. We achieve it by explicitly constructing a bidding strategy as a solution to an ordinary differential equation, and by proving its optimality for the intricate infinite-dimensional minimax problem underlying worst-case regret minimization. Our construction provides a systematic and computationally-tractable procedure for deriving minimax-optimal bidding strategies. When the value distribution is continuous, it yields a deterministic strategy that maps each value to a single bid. We also show that our minimax strategy significantly outperforms the uniform-bid-shading strategies advanced by prior work. Finally, we analyze the impact of the value distribution on the performance loss, and find that it decreases as the buyer's values become more dispersed.

SD57

Building B Level 3 B316

Data Analysis and Optimization in Social and Communication Networks

Invited Session

Social Media Analytics

Chair: Alexander Semenov, University of South Florida, Shalimar, FL, United States

1 - An Empirical Analysis on Systems of Nonlinear Equations for Social Network Influence Propagation Stability

Alexander Semenov, University of South Florida, Shalimar, FL, United States, Michael Hirsch, Panayote (Panos) Pardalos

How information spreads throughout a complex network can be modeled through influence propagation, Stability analysis, done through simulation of the influence propagation on the network over time, allows one to realize the final effects of the influence on all nodes of the network. In this research, we investigate various influence propagation models and the stability of the underlying complex network through the solving of systems of nonlinear equations, and compare computational results when using simulation.

2 - Coordinated Routing of Uavs and Refueling Trucks for Persistent Aerial Surveillance Using Milp

Donald McMann, University of Florida, Shalimar, FL, United States, Yongpei Guan, Alexander Semenov

This work presents a mixed-integer linear programming (MILP) formulation for coordinating fixed-wing unmanned aerial vehicles (UAVs) and refueling trucks to enable persistent surveillance over large border or coastal regions. UAVs are tasked with patrolling a grid-based coverage area, subject to battery/fuel limitations and telemetry range constraints. Ground vehicles provide mobile refueling support and must rendezvous with UAVs to recharge them during operations.

Our formulation captures UAV and truck movements through a time-expanded network, enforcing flow conservation, co-location constraints for refueling, and rolling coverage windows that ensure each grid cell is visited at least once within a predefined time horizon. The model also accounts for practical constraints such as maximum UAV flight range from a ground station and truck-only access nodes.

We implement the model in Pyomo solver and demonstrate its performance on simulated grid instances. The results highlight trade-offs between coverage persistence, battery constraints, and truck mobility. A Complexity analysis indicates that the problem is NP-hard, motivating future work on decomposition-based or heuristic approaches.

This research contributes to the growing body of work on autonomous aerial-ground system coordination and has direct applications in border security, coastal patrol, and post-disaster reconnaissance.

3 - TBA

Alexander Nikolaev, University at Buffalo, Buffalo, NY, United States

4 - Ultra-Small World Detection in Networks: Subgraphs with Prescribed Distance Distributions

Vladimir Boginski, University of Central Florida, Orlando, FL, United States

We introduce a class of distance-based clique relaxation models, which enforce certain distributions on vertex pairwise distances in the corresponding induced subgraphs. Both “local” and “global” versions of the proposed approach are studied. For the global case, the distribution of distances is considered for all vertex pairs in the subgraph. For the local (and, in a sense, more restrictive) case, the required property must be satisfied for any vertex with respect to its distances to the other vertices in the subgraph. We refer to the proposed clique relaxations as *global* and *local ultra-small worlds*, respectively. Our modeling approach has several meaningful interpretations; in particular, it is closely related to the concept of an “effective” graph diameter. Furthermore, density- and degree-based quasi-cliques are two well-known special cases of the proposed concept. We exploit these relationships to develop mixed integer programs (MIPs) that can be used with an off-the-shelf solver for finding maximum global and local ultra-small worlds. Then, we outline a simple-to-implement algorithm that iteratively solves feasibility versions of our MIPs for each possible subgraph size within some lower and upper bounds; we derive one non-trivial upper bound using the linear programming relaxations. Our modeling approach also generalizes the *k*-club concept; hence, some links to this well-known distance-based clique relaxation are also explored. Finally, to illustrate the obtained results, we perform a computational study on graphs representing various types of real-world datasets and systems. Some interesting empirical observations and insights are provided.

5 - The Effect of Repetition and Complexity Within and Between Topics on Social Media Engagement

Timothy Kaskela, Oregon State University, Corvallis, OR, United States, Jaeki Song

Social media allows organizations to communicate with individuals at a rapid pace. Research has traditionally examined the effect of a single post on engagement and other factors. However, there is an opportunity to examine the within topic effects of repetition and complexity as organizations post across longer periods of time, as the pace of communication on social media platforms allows an organization to quickly shift between multiple topics in a short time frame. This study examines the effect of repetition on engagement, as well as how complexity embedded within each post moderates said effect. A pilot study is conducted on the social media posts from 5 NASDAQ 100 companies to examine how repetition within and between posts impacts engagement, as well as how the complexity of the posts moderates this effect. The pilot study findings suggest that there is a u-shaped relationship between topic repetition and engagement, while post complexity has an inverted u-shaped direct effect on engagement. In addition, post complexity does indeed moderate the relationship between topic repetition and engagement. The findings suggest a full study examining the dynamics of repetition and complexity on social media is warranted.

SD58

Building B Level 3 B317

Analytics for Manufacturing

Contributed Session

Contributed

Chair: Luis Alcala Becerra, NORTH DAKOTA STATE UNIVERSITY, Fargo, ND, United States

1 - New Optimization Method for Assigning Vehicles to Transportation Products in Steelworks

Katsuki Shimada, JFE Steel Corporation, Fukuyama-shi, Japan, Shinji Tomiyama

In the large steelmaking companies, the products transportation between factories, warehouses, and shipping berths by special vehicles plays an important role. The vehicle assignment schedule for products transportation is made and revised manually for 24 hours. However, achieving efficient scheduling is quite challenging due to the complex interactions among many vehicles and cranes, which can be influenced by various disturbances. As a result, the current manual scheduling often leads to inaccuracies, causing increased vehicle waiting times and canceled deliveries.

In this study, we develop the new computer algorithm to generate vehicle assignment schedules for steel coil products transportation within steelworks. The proposed algorithm aims to generate feasible vehicle assignment schedules that can adapt to real-time changes of operations. It effectively distributes workloads for both vehicles and cranes by estimating the total workload for each facility. Additionally, by using a rule-based approach, the algorithm can generate a schedule in accordance with actual operations. Furthermore, the algorithm is capable of computing schedules in under one minute using straightforward logic.

To evaluate the performance of the proposed algorithm, we conducted simulation experiments. The results of simulation show that the algorithm can generate vehicle assignment schedules that reduce the number of transportation cases exceeding the scheduled completion time for transportation jobs by approximately 45% compared to manual scheduling.

The performance of the algorithm enables both efficient vehicle assignment scheduling and frequent rescheduling to adapt to the changes of operations.

2 - A Simulation-Based Sensitivity Analysis of Consensus Performance in Public Blockchains with Dynamic Committees

Jin Choi, Ajou University, Suwon, Korea, Republic of, Jaeyoung Lee

This study presents a simulation-based sensitivity analysis of consensus performance in public blockchains employing dynamic committees. In such systems, major nodes typically initiate the network, and consensus committees are formed dynamically by both major and minor nodes. We focus on node-level decentralization, defined as the proportion of a node's total stake that participates in the consensus process. Although this metric is intended to ensure fair participation, it can lead to centralization, as nodes with larger stakes tend to dominate committee involvement. To explore this issue, we conduct simulations by varying key parameters, including the number of major and minor nodes, average stake per node, stake distribution, and the number of consensus rounds. Our analysis reveals how these factors influence fairness and the risk of centralization in consensus participation. Based on the results, we propose an effective control strategy for managing dynamic committees, aiming to improve both fairness and efficiency in public blockchain networks.

3 - Developing Strategic Decision-Makers for the AI Era: Lessons from a Leadership Simulation Workshop

Joseph Cazier, Arizona State University, Tempe, AZ, United States, Daniel Mazzola, Gayatri Gade

As AI accelerates change across industries, the demands placed on future business leaders are evolving rapidly. Success now requires more than analytical ability—it depends on strategic decision-making, systems thinking, and the ability to lead under uncertainty. Yet many graduate programs struggle to deliver these skills in ways that reflect the complexity and pressure of real-world environments.

To address this gap, we designed and delivered an experiential leadership simulation for graduate business students. In this hands-on workshop, participants managed virtual companies across multiple competitive rounds, making decisions on pricing, R&D, hiring, and capital allocation. Market conditions changed dynamically, requiring teams to analyze data, respond to shifting variables, and collaborate under pressure.

The simulation was designed to align with key learning goals in critical thinking, communication, leadership, data-driven decision-making, and systems integration. Participants engaged in structured self-assessments and received individualized feedback on their decision-making processes and team leadership behaviors.

Results showed strong participant engagement and perceived skill development, with high satisfaction and demand for continued offerings. Qualitative feedback emphasized the value of learning through consequence-rich decisions in a low-risk setting—replicating the challenges of executive leadership in an AI-enabled economy.

We argue that simulation-based learning provides a scalable, high-impact model for graduate education. It allows students to move beyond theory and practice complex, integrated skills essential for leading in the digital age. We conclude with recommendations for integrating simulations into the core curriculum and for further research on longitudinal skill retention and performance transfer.

4 - A Stochastic Optimization Approach to Robust and Equitable Surgeon Workload Allocation

Jonathan Akhagbosu, University of Massachusetts Amherst, Amherst, MA, United States, Hari Balasubramanian, Tovv Kamine, Muge Capan

Surgeon workload influences clinical outcomes, provider well-being, and operational efficiency. Yet, existing research in surgical operations has largely focused on optimizing system-level resources such as operating room utilization or bed capacity, rather than considering the surgeon as a constrained resource. In addition, there is a need for incorporating workload equity in surgical case allocation to ensure optimal use of surgeons as resources. This study presents a stochastic integer programming model that incorporates workload equity into surgical case allocation. Workload is quantified using surrogate indicators, including surgery duration, American Society of Anesthesiologists (ASA) classification, case complexity, and procedure timing (regular vs. off-hours). Based on these indicators, surgical cases are grouped into discrete workload types using k-means clustering. This workload categorization informs the cost parameters used in the objective function. Equity is defined through a pairwise envy-based metric, which measures the absolute difference in total workload between any two surgeons. The model minimizes the maximum pairwise workload envy across all surgeon pairs, subject to constraints on surgical volume, surgeon availability, and uncertainty in surgery durations. Surgical cases are assigned to surgeons over a fixed planning horizon. The model was applied to a dataset of over 86,000 surgical cases performed between 2020 and 2023 across multiple specialties at a large academic medical center. Preliminary results indicate reduction in maximum pairwise workload envy by up to 87% relative to baseline assignment practices. Future research will expand the workload definition to include non-surgical tasks and develop equity-aware scheduling policies that preserve capacity for emergency cases.

5 - Payroll Minimization in Electronics Contract Manufacturing via Role-Specific Assignment of Direct, Indirect, and Managerial Labor

Luis Alcalá Becerra, North Dakota State University, Fargo, ND, United States, Diana Lopez Soto, Mohajid Saaed Osman

Electronics contract manufacturing (ECM) is a \$500-billion-dollar industry and is expected to grow to \$1,000 billion by 2030. This industry usually operates under tight resource constraints, a high product mix, and volatile demand. The industry's average operating margin is around 3%, with payroll costs representing approximately 10% of total revenue. These challenges make workforce planning a critical activity for proper operations. Current literature shows multi-skilled assignment for single departments, solely focusing on direct labor without considering the effect indirect labor has on payroll costs. This study presents a mixed-integer linear programming framework tailored to the ECM industry. The model incorporates 1) direct and indirect labor cost structures, 2) role-specific skills and productivity differences, 3) multi-department assignments, and 4) partial allocation of managerial personnel. The objective of the model is to minimize payroll expenses. The model is validated using multiple instances generated based on real scenarios. This framework provides valuable information to managers tasked with personnel reallocation and operating costs optimization on the ECM industry.

SD59

Building B Level 3 B318

Dynamic Pricing and Revenue Management

Contributed Session

Contributed

Chair: Sichen Guo, Shanghai University of Finance and Economics, Shanghai, N/A, China, People's Republic of

1 - Time-Sensitive Price Discount Optimization for Seller Facing Strategic Buyers with Stochastic Arrival

Suyeon Choi, KAIST, Daejeon, Korea, Republic of, Changhyun Kwon, Seungki Min

We investigate the optimal discount pricing schedule for a single indivisible perishable item over a finite sales horizon, during which strategic buyers arrive stochastically.

Due to the item's perishability, these buyers may delay their purchases in anticipation of more favorable prices, while risking the possibility of stockout.

To facilitate analytical tractability, we restrict each buyer's strategy to depend separately on their arrival time and private valuation.

This leads us to introduce a novel framework—Value-Arrival Separable strategies—which decouples the interdependence inherent in general strategies and enables the characterization of equilibrium behavior via an ordinary differential equation.

Building on this framework, we analyze the seller's efficient frontier, which identifies the optimal trade-off between expected revenue and expected buyer waiting time.

Specifically, we characterize near-optimal pricing schedules under two distinct market regimes.

In thick markets with high buyer arrival rates, the optimal schedule follows a linear discounting rule, with the slope determined by seller's time sensitivity.

In contrast, in thin markets with low arrival rates, a constant pricing schedule is optimal.

The shape of the optimal pricing schedule endogenously depends on the arrival rate of buyers.

Our findings contribute to the literature on dynamic pricing by providing practical insights for e-commerce and retail platforms managing expiring goods in the presence of strategic consumer behavior.

2 - Conditional Generative Learning for Joint Pricing and Inventory Control

Xianghua Jiang, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Xiangbin Hu, Xun Zhang, Zhisheng Ye

We propose a provable deep conditional generative framework for data-driven joint pricing and inventory control with contextual information, in which the retailer makes pricing and inventory decision based solely on historical data consisting of demand, price and covariates.

In contrast to the existing methods that directly learn the optimal policy curve as a function of the covariate, we develop a deep conditional generative learning approach to learn from historical data the whole demand distribution conditional on selling price and covariates.

The conditional demand estimator is a deep conditional generator, which is a function of the selling price, covariates and a 'noise' variable independent of price and covariates.

The special structure of the deep conditional generator allows us to sample the noise easily and compute the gradient of the random cost in closed form, with which stochastic gradient descent can be designed for the subsequent policy optimization under both risk neutral and risk averse settings.

We establish asymptotic properties of our data-driven policy including asymptotic optimality and consistency, and we provide a consistent estimator of the optimal expected cost.

Through comprehensive simulations, we show that the expected cost evaluated at our data-driven policy is significantly smaller than those from state-of-the-art prescriptive benchmarks.

In particular, we apply our method to real meal delivery data and find that our method beats the benchmarks by at least 10% reduction in out-of-sample costs, and it is at least four times faster than three machine learning-based descriptive approaches.

3 - Optimal Pricing in Trade-in Programs: the Consequences of Rebate Adjustments for Firm Profits and Social Welfare

Haoying Guo, Southeast University, Nanjing, China, People's Republic of, Yi Ding

As trade-in programs gain increasing popularity, firms have adopted various pricing strategies, including basic, estimated, and range-estimated pricing. The basic pricing strategy determines trade-in rebates immediately based on the condition of the used product. The estimated pricing strategy operates in two stages: an initial rebate estimate is offered, followed by a final trade-in rebate upon receipt of the used item. In contrast, the range-estimated pricing strategy initially sets a rebate range rather than a fixed amount. We develop game-theoretical models to compare these three strategies. Our analysis yields several key insights. First, when firms adopt the estimated pricing strategy, they tend to lower final trade-in rebates, increasing their profits at the expense of replacement consumers' surplus. Notably, this reduction in trade-in rebates may also negatively impact social welfare. Second, under range-estimated pricing, consumers form precise psychological expectations about trade-in rebates, which reduces trade-in demand. In response, firms optimally lower the selling price of new products, thereby improving consumer surplus for new consumers. Finally, the range-estimated pricing strategy does not universally dominate the basic or estimated pricing strategies in terms of profits or social welfare. Their relative performance depends on market size, consumer sensitivity, and the durability of used products.

4 - Dynamic Pricing of Split Stays

Yi Zhuang, Zicklin School of Business, Baruch College, CUNY, New York, NY, United States, Arash Asadpour, Yuan-Mao Kao

Traditional hotel stays are typically based on fixed check-in and check-out dates, with guests occupying the same room throughout their stay. In 2022, Airbnb introduced a "split stay" feature, allowing customers to book separate accommodations within a single trip. In this work, we study optimal pricing strategies for split stays in the hotel and accommodations industry, where a split stay refers to a guest switching rooms within the same hotel over the course of a trip. We model the seller's decision problem as a dynamic program, where prices for both standard and split-stay options are adjusted over a finite selling horizon based on time and room availability. Under the optimal policy, split stays are used to balance inventory when room availability is uneven across nights within the hotel. The timing of when split stays are offered is also crucial. Early in the selling horizon, the seller anticipates high future demand and conserves inventory for standard high-value bookings. Closer to the end of the horizon, the seller becomes more myopic, favoring high-value bookings that yield immediate revenue. As a result, the mid-horizon emerges as the optimal window for offering split stays—when there is neither strong demand certainty nor urgent revenue pressure. In this period, split stays are most effective for smoothing inventory use and maximizing overall revenue.

5 - Relatively Robust Multiproduct Pricing

Ru Zhang, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Thomas Weber

This study presents a relatively robust pricing framework for multiproduct monopolies under distribution-free demand uncertainty characterized by Frobenius-norm constraints. Compared to extant models based on interval uncertainty, our approach employs ambiguity sets that align naturally with standard multivariate estimation techniques. The optimal robust price vector, obtained as the solution to a quasi-concave programming problem, maximizes the worst-case ratio of achieved profit to the ex-post optimal profit under full information. A key structural result is that the performance index is strictly quasi-concave in price, allowing the unique optimum to be characterized by a first-order condition. An embedded fixed-point problem can be solved efficiently, and closed-form expressions are derived for symmetric linear, semi-log, and log-log demand systems with associated parameter uncertainty. In general nonlinear settings, the model remains applicable through local linearization and regression-based confidence regions. To evaluate the approach, we conduct both extensive simulation experiments and empirical analyses using large-scale datasets from a major online retailer. The results demonstrate the superior performance of the relatively robust pricing method compared to alternative robustness criteria such as maximin payoff, minimax regret, and Laplacian uncertainty.

6 - Switching Gradient: Learning to Compete with Price Matching Guarantee

Sichen Guo, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, Shukai Li, Cong Shi, Chaolin Yang

We study a sequential pricing problem in a duopoly market, where two sellers compete through dynamic pricing to maximize individual revenues without prior knowledge of their respective demand functions. We consider a setting in which one seller offers a price-matching guarantee (PMG), committing to match the competitor's price upon customer request. We show that, under positive hassle costs, the introduction of PMG fundamentally alters the structure of the revenue function, introducing non-convexities and discontinuities. These structural complexities render conventional learning algorithms and analytical tools inapplicable. To address these challenges, we propose a switching gradient framework capable of handling piecewise-quadratic revenue functions with discontinuities. We prove that our algorithm achieves a last-iterate convergence rate of $O(\log T / \sqrt{T})$ and a dynamic regret bound of $O(\log T \sqrt{T})$, relative to the optimal dynamic benchmark.

SD60

Building B Level 3 B319

AI Research and Innovation

Contributed Session

Contributed

Chair: Sunil Erevelles, University of North Carolina, Charlotte, 11501 Willows Wisp Dr, Charlotte, NC, 28277-8400, United States

1 - Interpretable Low-Rank Discriminative Cvae for Fault Clustering

Nan Li, Tsinghua University, Beijing, China, People's Republic of, Chen Wang

Data-driven models have become foundational in system health monitoring, enabling automated analysis of complex operational data. However, the inherent opacity of many advanced techniques limits their deployment in real-world environments where interpretability is critical. To address this challenge, we propose a novel Bayesian modeling framework that enhances interpretability by explicitly disentangling operational context—extraneous factors unrelated to system health—from sensor observations. Our approach integrates a conditional variational autoencoder (CVAE) with a mixture-based latent prior, further augmented by a low-rank discriminative regularization. This design encourages the extraction of low-dimensional, structured latent representations that promote meaningful clustering and facilitate intuitive interpretation of fault patterns. Beyond latent space modeling, we introduce a covariance-based interpretability module that systematically analyzes reconstructed sensor

data. This module identifies dominant sensor patterns associated with failure modes and degradation trends, offering a data-driven pathway for unsupervised fault clustering and fault characterization.

We validate the proposed approach through comprehensive experiments on benchmark datasets and operational records from service systems. Results demonstrate the framework's effectiveness in isolating context-independent health states, uncovering latent fault structures, and delivering actionable insights for fault diagnosis and predictive maintenance. Our work bridge the gap between data-driven models and the practical needs of transparent decision-making in health management applications.

2 - Optimizing Rag System Configurations Through Task-Specific Search Strategies

Yajuan Wang, Shanghai JiaoTong University, Shanghai , China, People's Republic of, Lai Wei

Retrieval-Augmented Generation (RAG) systems have become an effective approach to addressing hallucination and improving factual accuracy in Large Language Models (LLMs), especially in knowledge-intensive applications. However, most existing RAG implementations adopt a fixed, one-size-fits-all configuration across different tasks and datasets, overlooking the considerable differences in retrieval requirements, generation objectives, and domain complexity. This uniform design limits the adaptability of the system and reduces the overall quality of generated outputs in real-world scenarios. In this study, we present a task-adaptive RAG optimization framework that systematically identifies suitable configurations for different types of tasks. Using a network-based search method, we examine combinations of key RAG components—such as retrievers, top-k values, and reranking options—across two distinct datasets: Natural Questions and PubMedQA. These tasks differ significantly in their factual requirements, domain specificity, and tolerance for inaccuracy, making them well-suited for analyzing the impact of configuration choices. Our results show that task-specific configurations consistently outperform fixed, uniform baselines in terms of METEOR, ROUGE, and Semantic Similarity Scores. In addition, we identify clear patterns linking task characteristics with optimal configuration settings, offering both theoretical insights and practical guidance for the deployment of intelligent systems. This study contributes to the field of AI system management by proposing a scalable, data-driven approach to configuration adaptation in generative architectures.

3 - The New Knowledge Worker: AI as An Active Contributor to Organizational Knowledge Management

Matthias Bickel, ETH Zurich, Zurich, Switzerland, Torbjørn Netland

How does Artificial Intelligence change the way knowledge is created and shared? For decades, Nonaka's SECI model has been a foundational framework in knowledge management, describing how tacit and explicit knowledge are transformed through four stages: Socialization, Externalization, Combination, and Internalization. However, as AI becomes increasingly integrated into how individuals and organizations interact, collaborate, and make decisions, this model should be reconsidered. In this research, we revisit the SECI model and propose an updated version that reflects AI's evolving role—not just as a tool, but as an active participant in knowledge processes. We examine how specific AI technologies contribute to this shift: Generative AI and Inverse Reinforcement Learning help capture and formalize tacit knowledge during the externalization phase, while Reinforcement Learning supports the synthesis and integration of explicit knowledge in the combination phase. To demonstrate the practical relevance of the revised framework, we apply it to the airline industry. For this, we focus on operational challenges such as disruption management, tail assignment, and crew rostering. These examples show how AI technologies can improve the utilization and effectiveness of knowledge use in operational decision-making. We contribute a new perspective on knowledge management in the context of AI and offer practical implications for organizations implementing AI in operational contexts.

4 - Generative AI-Enhanced Mathematical Programming Workshop with Fico Xpress

Alkiviadis Vazacopoulos, Optimization Direct, Inc., Harrington Pk, NJ, United States

This workshop introduces participants to the transformative potential of generative AI in mathematical programming using FICO Xpress. Through hands-on demonstrations and practical exercises, attendees will learn how to leverage generative AI to streamline model development, enhance code generation, and improve result interpretation. The workshop covers key aspects including automated model formulation, intelligent debugging assistance, and AI-powered console output analysis. Participants will gain practical experience in using generative AI to solve complex optimization problems more efficiently, interpret solver outputs more effectively, and generate comprehensive documentation. This workshop is designed for optimization practitioners, researchers, and developers who want to enhance their mathematical programming workflow with state-of-the-art AI capabilities.

5 - Speeding up Grover's Algorithm

Stefan Creemers, UCLouvain, Louvain-la-Neuve, Belgium

To find one of the M target entries in an unstructured database that has N entries, Grover's algorithm performs $\pi/4(N/M)^{1/2}$ iterations. In each iteration, a function is called that evaluates whether an entry is a target entry. Compared to a classical procedure, that requires N/M evaluations, Grover's algorithm achieves a quadratic speedup. This quadratic speedup, combined with its general applicability as a search algorithm, have made Grover's algorithms one of the most important algorithms available today. A study to improve its runtime is more than justified. For this purpose, we investigate two speedup strategies: (1) reducing the number of iterations that are performed in each run of Grover's algorithm and (2) partitioning the database. For each of these strategies, we not only execute Grover's algorithm in parallel, but also explore the possibility to execute Grover's algorithm in series on a single Quantum Processing Unit (QPU) as well as on multiple QPUs. For each combination of execution mode (serial, parallel, and serial/parallel) and speedup strategy, we show how to obtain optimal policies using closed-form expressions and a gradient-descent procedure. For a single QPU we obtain a speedup factor of 1.1382 when compared to a textbook implementation of Grover's algorithm. If multiple QPUs are at our disposal, we obtain a speedup factor of at most $1.1382 Q^{1/2}$, where Q denotes the number of QPUs. In addition, we show that the dominant policies that minimize the expected number of Grover iterations also minimize the expected number of iterations that are performed per qubit.

6 - Operationalizing Artificial Intelligence in Pharmaceutical R&D: from Strategy to Scalable Impact

Shujaullah Khan Mohammed, AstraZeneca, Washington DC, DC, United States

Artificial Intelligence (AI) is rapidly reshaping the pharmaceutical R&D landscape, offering unprecedented opportunities to improve efficiency, accuracy, and speed across the value chain. Yet many organizations struggle to transition from fragmented pilots to enterprise-scale impact. This paper presents a comprehensive framework for operationalizing AI across core R&D functions including clinical operations, Chemistry, Manufacturing & Controls (CMC), medical writing, and real-world evidence (RWE) analysis.

A five-stage AI maturity model is introduced, ranging from experimental deployments to agentic AI systems capable of autonomous decision support. Detailed use cases demonstrate AI's potential to optimize clinical trial protocols, enhance patient recruitment, automate site monitoring, predict manufacturing stability, generate regulatory documentation, and extract insights from complex omics and real-world datasets. The paper highlights the growing role of AI agents in performing tasks traditionally handled by data scientists and medical writers, allowing organizations to redeploy human expertise toward more strategic and creative work.

A four-phase implementation roadmap is outlined covering strategic alignment, platform and data modernization, value realization, and enterprise scaling. Drawing on recent academic publications, industry benchmarks, and regulatory trends, this paper provides actionable guidance for pharmaceutical R&D leaders seeking to embed AI as a core operating capability. The future of drug development lies in systems that can reason, learn, and collaborate, transforming AI from a supportive technology into a strategic partner across the lifecycle of therapeutic innovation.

7 - Generative Artificial Intelligence and the Primacy of Public Policy: Using Self-Evolving Algorithms to Regulate an Algorithmic Superspecies

Sunil Erevelles, University of North Carolina, Charlotte, Charlotte, NC, United States, Anjali Sinha, Sahar Maleki, Emanuel Valenzuela

Generative artificial intelligence (AI) differs from *all* past technologies in that it is not a tool, but has independent agency. A tool can be controlled by humans and gives humans more power. Generative AI, by contrast, has the capacity to *create, decide, and execute* on its own. It has higher emotional intelligence than humans, specifically because it is emotionless. It turns humans into algorithmic inputs, and is increasingly able to precisely understand and control us. It evolves faster than humans, making it impossible for humans to catch up. In conformance with the biological and ecological species concepts, generative AI is fast exceeding the accepted criteria for species,

but inorganically. Accordingly, we define generative AI as an "inorganic superspecies." For the first time in human history, a species superior to humans is emerging. While the benefits of generative AI to humanity are remarkable, it also is likely that generative AI poses considerable danger to humanity, if not managed *proactively*. The purpose of this research is to help develop self-regulating mechanisms for generative AI public policy. Methodologically, the authors use transformer-based attention mechanisms, as well as an understanding of traditional support vector machine models, additive explainability analysis and adaptive moment estimation algorithms to create an inductive realist-based framework for regulating generative AI. This initial framework, consisting of foundational premises based on generative AI logic, could help contribute to academic advances in self-regulating generative AI techniques, as well as help practitioners use generative AI in a more effective and safe manner.

SD61

Building B Level 4 B401

Advanced Data-Driven Insights for Complex Industrial Processes

Invited Session

Quality, Statistics and Reliability

Chair: Zimo Wang, SUNY Binghamton, Binghamton, United States

Co-Chair: Chao Wang, University of Iowa, Iowa City, IA, United States

1 - Generating Synthetic Aia Images from Historical Ground-Based Observations: a Diffusion Model Approach for Extended Coronal Hole Analysis

Marco Marena, New Jersey Institute of Technology, Newark, NJ, United States, Bo Shen, Qin Li, Haimin Wang

Accurate and continuous observations of the Sun's extreme ultraviolet (EUV) emissions are critical for understanding solar dynamics and monitoring solar activity. However, EUV observations availability is limited to the past ~30 years, restricting our ability to study long-term solar corona. This study aims to reconstruct a historical dataset of synthetic AIA 193 Å and 304 Å — solar corona and chromosphere — using historical ground-based chromospheric observations in H α and CaII K, available for over a century. To generate AIA observations before 2010, we propose to use a novel deep learning methodology, the Diffusion Model Translation (DMT) framework, built on a fine-tuned Denoising Diffusion Probabilistic Models (DDPM), and a Two-Stream Image-to-Image Translation (TSIT) network as a translator between the noising and denoising processes, to bypass the diffusion steps and accelerate training. We used H α and CaII K lines from Kanzelhöhe Solar Observatory (KSO) spanning from 2011 to 2024, and AIA 193 Å and 304 Å from the Solar Dynamic Observatory (SDO). For the independent test set, the model shows a high full-disk correlation (CC = 0.942) and achieves a Dice coefficient of 0.893 for coronal holes. Particularly, the high Dice coefficient highlight the model's effectiveness in identifying and delineating coronal holes, sources of high-speed solar wind streams. By providing reliable synthetic EUV imagery for historical periods lacking direct observations, our methodology facilitates deeper analyses of coronal hole evolution, as well as long-term solar activity trends.

2 - An Advanced Spectral Shapley-Based Active Learning to Analyze Extreme Weak Sensor Signals for Precision Process Monitoring

Xinchen Wang, SUNY Binghamton, Binghamton, NY, United States, Zimo Wang

Understanding streaming sensor signals is essential for in-process monitoring and quality inspection for various industry applications. These signals contain rich, time-sensitive information that supports real-time detection of anomalies and quality deviations. However, at the micro- and nano-scale monitoring process, the acquired signals are often overwhelmed by high levels of noise and interferences with unknown structures, which may significantly impact the accuracy and effectiveness of signal interpretations where conventional filtering methods—such as bandpass filtering or denoising autoencoder—may be insufficient for isolating relevant information. To address these challenges, we introduce a Hierarchically Influenced Spectral Shapley-based Active Learning (HISS-AL) approach, which learns the underlying information from streaming sensor data. The method is first validated using music timbre recognition and then applied to real-time surface morphology characterization during nanomachining. Results confirm the model's ability to capture intrinsic spectral features in simulated and real acoustic environments. In nanomachining, the framework achieves an overall accuracy of 84% for predicting surface morphology based on the signal gathered during the nano-scale material removals. These findings highlight the method's robustness in high-noise, low-SNR environments, where conventional approaches often falter. Overall, the proposed strategy offers a promising direction for advancing sensor-based monitoring and quality control in precision manufacturing.

3 - Physics-Informed Pipeline Corrosion Reliability Prediction Using Inspection Images

Nikesh kumar, University of Arkansas, fayenteville, AR, United States, miao tan, Haitao Liao

This study presents a hybrid framework for predicting the reliability of pipelines subject to internal corrosion by integrating image-based defect extraction, finite element stress analysis, and stochastic degradation modeling. Using machine learning methods, corrosion geometries including depth, length, and width are extracted from inline inspection images, even in challenging areas such as welds and branches. These geometric features drive the development of a physics-informed stochastic model. Pit initiation is represented by a nonhomogeneous Poisson process, where the rate of new pit formation varies with time and environmental conditions such as temperature, pressure, and fluid chemistry. The depth of each pit evolves over time according to a Gamma process, while the defect length and width grow with linear trends. The proposed formulation allows the model to capture the full three-dimensional evolution of corrosion. To manage computational complexity, finite element simulations are conducted to first generate a library of stress fields for varying defect shapes. These are used to scale growth rates according to local stress intensity. Instead of simulating each defect individually, the model interpolates stress effects based on the matched geometries from inline images. Monte Carlo simulation samples random environmental conditions and material properties, producing failure probabilities over time by tracking whether cumulative damage surpasses critical thresholds.

Validation is performed using offshore pipeline data. The model's predictions closely match historical failure trends, demonstrating accuracy in assessing pipeline failure risk. Compliant with ASME B31G and API 579 standards, the framework enables proactive, image-guided maintenance planning and supports scalable pipeline integrity management.

4 - Diffusion Models for Anomaly Detection in General Textured Surfaces

Mehrdad Moradi, Georgia Institute of Technology, Atlanta, GA, United States, Kamran Paynabar, Marco Grasso, Bianca Maria Colosimo

Anomaly detection at the pixel and instance level in textured patterns has gained considerable attention, with both statistical and machine learning approaches demonstrating strong performance. However, most existing methods are limited to modeling a **single pattern type**, restricting their applicability in complex or multi-pattern environments. While neural networks offer greater flexibility by learning multiple patterns within a single model, they often suffer from limited interpretability and require large volumes of labeled data.

In this work, we introduce a novel diffusion-based methodology for anomaly detection that is theoretically grounded and capable of handling **multiple textured patterns** without prior knowledge of their number or form. Our approach partitions input data into patches and applies a newly developed **pixel-level anomaly detection mechanism** based on reverse diffusion dynamics. Experimental results show that our method outperforms both state-of-the-art diffusion-based models and traditional statistical techniques.

5 - Multivariate Nonparametric Functional Calibration of Wake Effect Models

Cheoljoon Jeong, University of Michigan, Ann Arbor, MI, United States, Eunshin Byon

As wind energy continues to grow, understanding wake effects—wind shadows cast by upstream turbines that reduce downstream power output—has become increasingly important. Accurately modeling these effects requires calibrating parameters within simulation models using operational data. Prior studies often fail to capture the complex dependence of parameters on multiple environmental factors. This study proposes a new nonparametric functional calibration method that estimates parameters without assuming a predefined functional form of the parameters. Our method accounts for multivariate environmental influences and is validated through both simulated and real-world wind farm case studies.

SD62

Building B Level 4 B402

QSR Best Student Paper Competition

Award Session

Quality, Statistics and Reliability

Chair: Bo Shen, New Jersey Institute of Technology, Newark, United States

Co-Chair: Cesar Ruiz, University of Oklahoma, NORMAN, United States

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

1 - Locally Transparent Rule Sets for Interpretable Machine Learning**Brian Liu, MIT, Cambridge, MA, United States, Rahul Mazumder**

We introduce the concept of local transparency in machine learning: models in which predictions on a subset of the data can be explained concisely. We propose an optimization-based framework for constructing locally transparent rule sets and develop algorithms to efficiently evaluate the trade-off between local transparency and accuracy.

2 - Natural Hypergradient Descent: Algorithm Design, Convergence Analysis and Parallel Implementation**Deyi Kong, University of Minnesota, Minneapolis, MN, United States, Zaiwei Chen, Shuzhong Zhang, Shancong Mou**

Bi-level optimization has emerged as a powerful framework for modern machine learning applications such as meta-learning, data synthesis, and hyperparameter tuning. Hypergradient descent is commonly used to solve bi-level optimization problems, but computing the hypergradient—particularly the Hessian inverse of the inner problem, either explicitly or implicitly—poses significant computational challenges when the inner problem involves large-scale models. In this paper, we propose a simple and scalable method, Natural Hypergradient Descent (NHGD), which follows an optimize-and-approximate strategy. NHGD approximates the Hessian inverse using a running average of the inverse empirical Fisher information matrix (EFIM), computed directly from the gradients generated by each inner-loop SGD update. We provide a detailed non-asymptotic analysis demonstrating convergence to stationary points with iteration complexity matching that of existing optimize-then-approximate hypergradient methods, which first solve the inner problem to a desired accuracy and then approximate the Hessian inverse via Neumann series or conjugate gradient solvers. A key advantage of NHGD is that its optimize-and-approximate design enables parallel computation of the hypergradient approximation alongside the inner-loop SGD steps. This leads to substantial improvements in computational efficiency, particularly for large-scale problems. Experiments on hyper data cleaning and data distillation tasks demonstrate that NHGD achieves comparable or better performance.

3 - Theoretical Analysis for Optimal Design of An Online Monitoring and Sampling Scheme Under Partial Observation**Dongmin Li, Georgia Institute of Technology, Atlanta, GA, United States, Xiaochen Xian**

In modern quality control, online monitoring of high-dimensional data streams has become increasingly important as large volumes of real-time data are collected from diverse sources. However, practical constraints such as limited sensors, transmission bandwidth, and storage capacity often restrict data availability to only a subset of variables at each time. In such resource-constrained settings, the determination of key parameters, particularly the number of observed variables, is critical to balance detection capability with resource costs to achieve optimal statistical and/or economic performance. This parameter design relies on analyzing the relationship between detection delay and parameter values, but such analysis is lacking in existing monitoring schemes under partial observations. While some studies provide upper bounds for average detection delay, these bounds are either asymptotic or excessively large, offering limited practical guidance. To address this gap, we develop an optimal design framework based on theoretical analysis of detection delay for a monitoring-based probabilistic sampling scheme that adaptively selects observed variables at each time. Leveraging sampling probabilities that are explicit functions of monitoring statistics, we derive formulas for average detection delay in finite settings as functions of design parameters and system settings, such as the number of observed and total variables. Enabled by the theoretical results, we propose an economic-statistical design framework that determines the optimal number of observed variables to meet detection goals while accounting for practical costs. Simulation studies across various parameters and system settings validate the effectiveness of the theoretical results and design framework, demonstrating significant practical value for optimal design.

4 - Tensor-Efficient High Dimensional Q-Learning**Junyi Wu, University of Washington, Seattle, WA, United States**

High-dimensional reinforcement learning faces challenges with complex calculations and low sample efficiency in large state-action spaces. Q-learning algorithms struggle particularly with the curse of dimensionality, where the number of state-action pairs grows exponentially with problem size. Current solutions include function approximation methods and neural network-based approaches like Deep Q-Networks, but these often suffer from instability and require extensive hyperparameter tuning. We propose Tensor-Efficient Q-Learning (TEQL), which applies low-rank tensor decomposition via block coordinate descent on discretized state-action spaces, incorporating novel exploration and regularization mechanisms. The key innovation is an exploration strategy that combines approximation error with visit count-based upper confidence bound to prioritize actions with high uncertainty, avoiding wasteful random exploration. Additionally, we incorporate a frequency-based penalty term in the objective function to encourage exploration of less-visited state-action pairs and reduce overfitting to frequently visited regions. Empirical results on large discrete control tasks demonstrate that TEQL outperforms conventional matrix-based methods and deep RL approaches in both sample efficiency and total rewards, making it suitable for resource-constrained applications, such as space and healthcare where sampling costs are high.

SD63

Building B Level 4 B403

Forecasting and Uncertainty Quantification

Contributed Session

Contributed

Chair: Seyed Soroosh Alavi, Auburn University, 313 E. Thach Ave., Auburn, AL, 36830, United States

1 - A Discrete-Time Markov Chain for Lifetime of Lithium-Ion Batteries**Jeongeom Lee, Seoul National University of Science and Technology, Seoul, Korea, Republic of, Min Kyu Sim**

Lithium-ion batteries (LiBs) are pivotal for energy storage systems (ESS) and electric vehicles (EVs). Their capacity, represented by the State of Health (SoH), degrades over their service life. Because each use cycle incurs economic depreciation, understanding this degradation behavior is critical for devising optimal operation strategies. Most existing models divide battery degradation into two distinct phases: a “normal degradation” phase until the SoH reaches 80%, and a “knee curve degradation” phase characterized by a drastic decline after the SoH falls below 80%. This bifurcated approach makes it challenging to accurately and reliably estimate the battery’s Remaining Useful Life (RUL). To address this limitation, we propose a discrete-time Markov chain (DTMC) that integrates both phases. Specifically, each state in our model encodes a discretized SoH level and a binary knee-phase flag, enabling us to capture both degradation patterns simultaneously. Because the DTMC can serve as a stochastic building block, it offers various advantages. For example, simulation studies can yield not only point estimates but also interval estimates, allowing for risk-adjusted pricing. Moreover, the model can be used to analytically and numerically estimate a wide range of statistics, including means, confidence intervals, and complete distributions. Through empirical studies, we confirm that our model aligns well with existing publicly available datasets. From a practical perspective, it is expected to provide a fundamental guideline for second-life battery (SLB) applications in the near future.

2 - Latent Variable Transfer Learning for Modeling Energy Demand in Industrial Systems**Afshin Asadi, UNIVERSITY OF MIAMI, Miami, FL, United States, Ramin Moghaddass**

Accurate and adaptive energy consumption modeling is essential for optimizing industrial efficiency, reducing operational costs, and supporting grid stability. However, industrial systems pose signifi-

cant modeling challenges due to complex operational dynamics and latent factors such as equipment degradation, variable production schedules, and environmental fluctuations. Traditional models often lack the flexibility to generalize across different systems, especially when historical data is scarce. This paper presents a recurrent variational neural network (VRNN) framework that learns compact latent representations from historical energy data, capturing underlying patterns shared across industrial processes. Leveraging transfer learning, the proposed approach enables efficient adaptation to new systems by reusing learned representations and fine-tuning only the output layers. Numerical experiments demonstrate that this strategy improves prediction accuracy while preserving key system dynamics. The framework offers a data-efficient solution for industrial energy forecasting, highlighting the value of latent variable modeling and transfer learning in real-world energy management applications.

3 - Advances in Robust Linear Regression with Doubly Censored Data

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

This study evaluates a modification of median-based robust linear regression estimators to accommodate “general” censoring, where both the predictor and response variables can be left- or right-censored. We examine statistical properties of the resulting estimator and modifications to efficient algorithms for computing slope and intercept estimators that account for the censored observations.

4 - Uncertainty Quantification in Ros2 Gazebo Simulations: a Leader-Follower use-Case

Trevor McClain, University of Alabama in Huntsville, Huntsville, AL, United States, Rahul Bhadani

Robotics physics simulators are widely used to test the safety and correctness of autonomous robot control systems. In this work, we quantify the uncertainty present in the robotics control simulation environment, Gazebo Humble, and show that this uncertainty can lead to erroneous results on basic problem statements. Reproducibility in autonomous robot control systems simulations is of paramount importance to allow for useful data collection. The robotics control problem chosen for this work was the Intelligent Driver Model (IDM). In our setup, a fixed-velocity leader robot was followed by up to twelve follower robots. Our research showed that the Gazebo multi-robot simulation did not exhibit repeatable, or deterministic, behavior when the number of robots was seven or greater for the IDM car following model. This coincided with a degradation of the real time factor (RTF) of the simulations. Consistency in simulation results are vitally important for safety verification and data reproducibility, and our results raise concerns for the reproducibility and safety verification of simulation-based environments.

5 - Shift-Adaptive Multi-Step Conformal Prediction for Non-Stationary Time-Series

Marzieh Amiri Shahbazi, Rochester Institute of Technology, Rochester, NY, 14623, NY, United States, Ali Baheri

Shift-Adaptive Multi-Step Conformal Prediction, or SAM-CP, tackles the persistent challenge of producing reliable forecast uncertainty when both the mean and the variance of a time-series evolve over time. Operating fully online, the method first applies a lightweight cumulative-sum test to a sliding window of standardized residuals to sense distributional change. It then nudges a horizon-specific significance schedule through a Robbins–Monro update so that any empirical coverage shortfall is driven toward zero, and finally forms conformal prediction sets with the recalibrated non-conformity scores. We prove that this interplay of change-point monitoring and adaptive significance guarantees marginal coverage at the nominal level for every forecast horizon under arbitrary piece-wise stationary shifts, extending adaptive-conformal validity into the multi-step domain. Empirical studies on electricity demand, traffic flow, and macroeconomic benchmarks—each sliced into deliberately adversarial train-test regimes—show that SAM-CP maintains near-target coverage while producing substantially narrower prediction bands than baselines that rely on a fixed significance level or static residual weighting. Because the procedure wraps any existing forecaster without retraining, it offers practitioners a plug-and-play uncertainty layer that tracks evolving dynamics in real time, making it attractive for applications such as staffing, inventory, and energy dispatch.

6 - Forecasting Seasonal Changeoint Models Utilizing Changeoint Uncertainties

Vasileios Pavlopoulos, The University of Alabama in Huntsville, Huntsville, AL, United States, Rebecca Killick, Ege Can, Pham Hieu, Paras Bhatt

Forecasting in the presence of changeoints is a challenging task. However, if changeoints occur regularly across seasons, we want to utilize that information in our predictions for the next season. In this paper, we develop a framework for forecasting seasonal data that contain within-season changeoints. In our forecasts, we account for uncertainty in the changeoint locations both within and across seasons via a weighting approach across changeoint uncertainties. The framework is very flexible across different changeoint model assumptions and approaches to identifying changeoint confidence sets. We demonstrate the improvement in forecasting performance of our framework for 1) atmospheric pressure with a multivariate changeoint approach and 2) air pollution with a univariate changeoint approach.

7 - Revolutionizing Reliability Forecasting: Machine Learning vs. Finite Element Method and Empirical Models for Solder Joints

Seyed Soroosh Alavi, Auburn University, Auburn, AL, United States, Daniel Silva

Reliable estimation of solder joint lifetime is essential for maintaining the integrity of electronic assemblies, particularly when exposed to harsh environmental conditions. This work presents a comparative analysis of machine learning (ML) techniques alongside traditional power-law-based empirical models and finite element method (FEM) simulations carried out in ANSYS. Using an extensive dataset of 185 experimental tests, the study evaluates each method's performance. The ML framework integrates a wide range of isothermal aging and thermal cycling scenarios into a single predictive model, unlike empirical and FEM methods that require reconfiguration or recalibration for each specific condition and demand significant computational effort.

While FEM simulations offered valuable physical understanding, they were constrained by high computational costs and complexity in model setup. Empirical approaches, on the other hand, tended to lose accuracy under elevated temperatures or intricate component geometries. In contrast, ML models—particularly XGBoost and artificial neural networks (ANN)—demonstrated enhanced precision, scalability, and adaptability. These models effectively captured nonlinear relationships among both continuous (e.g., aging time, temperature) and categorical (e.g., surface finish, alloy type) variables. Evaluation metrics showed that ML predictions maintained high accuracy across all test conditions, with typical errors below 3%, outperforming both empirical and FEM strategies.

Overall, this study highlights the advantages of ML in predicting solder joint reliability, offering reduced computational requirements, simplified modeling processes, and improved predictive accuracy.

SD64

Building B Level 4 B404

Emerging Topics in Behavioral Operations Management

Invited Session

Behavioral Operations Management

Chair: Swanand Kulkarni, Tuck School of Business, Dartmouth College, Hanover, NH, United States

1 - Discretionary Task Switching and Knowledge Worker Productivity

Samer Charbaji, New York Institute of Technology - Manhattan, New York City, NY, United States, Roman Kapuscinski, Stephen Leider

Knowledge workers are often given creative tasks and repetitive tasks to work on and can have some discretion in how they switch between them. Research has shown that task switching lowers performance on repetitive tasks, but can improve performance on creative tasks. It is unclear if and how knowledge workers assigned both tasks would switch between them and what impact that would have on their overall performance. Our paper examines this using a laboratory experiment that incentivizes participants based on their performance on a creative task and a repetitive task and that gives them the discretion to freely switch between them. We first replicate the effect of task switching in our experimental setting. We run one treatment that forces participants to continuously switch tasks and one treatment that forces participants to

not switch tasks and show that task switching does indeed increase (decrease) performance on the creative (repetitive) task. We then examine participant behavior in the discretionary switch treatment and find that participants rarely switch tasks. This infrequent task switching results in a similar decrease in repetitive task performance as in the forced switch treatment, but, interestingly, without the increase in creative task performance, resulting in a significantly worse overall performance. We show that adding time-based and/or performance based nudges that encourage participants to switch tasks more often are surprisingly effective at encouraging participants to do so. This results in participants having a creative task, repetitive task, and an overall performance that is similar to the forced switch treatment.

2 - Going Local: Authenticity of Retailers' Buy-from-Local-Suppliers Programs

Orchi Bhattacharyya, University of Minnesota, Carlson School of Management, Minneapolis, MN, United States, Necati Ertekin, Karen Donohue, Mehmet Gumus

This study investigates Buy-from-Local-Suppliers programs, in which grocery retailers offer local food products in their assortment. Using data from a Canadian online retailer and an online experiment, we explore the impact of local certifications on sales of local food products, offering insights for effectively leveraging certifications in grocery retailing.

3 - Pay Model Intuitiveness and Transparency on on-Demand Service Platforms

Swanand Kulkarni, Tuck School of Business, Dartmouth College, Hanover, NH, United States, Basak Kalkanci, Chris Parker

On-demand service platforms have been experimenting with models that determine compensation for their workers. While some use commission- or effort-based pay models that are intuitive to workers, others, in their efforts to better match customer demand, have transitioned to models where pay is not strictly tied to effort, but depends on other, potentially exogenous factors. Platforms have also kept these pay models opaque. Workers' reactions to such practices, however, are not systematically examined or understood. Through incentivized experiments, we examine how a pay model's intuitiveness to workers, its transparency, and a change that reduces the model's intuitiveness affect workers' participation on the platform. Results show the promise and limits of pay model intuitiveness and transparency in managing worker participation.

4 - Does Experience Shape Online Reviews? the Role of Platform Interventions

Nalin Shani, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Achal Bassamboo, Maria Ibanez

This paper investigates how user experience and platform design jointly shape the perceived helpfulness of online reviews. We begin by examining how a reviewer's experience—measured via playtime—affects the structure, content, and perceived helpfulness of their review. While more experience might intuitively signal greater credibility, we find that this relationship is non-monotonic and differs for positive and negative reviews. These patterns are driven in part by how experience alters the emotional tone of the review, particularly along valence and arousal dimensions.

We then study two platform interventions tied to consumer experience: review solicitation and refund policies. We leverage two natural experiments generated by platform interventions tied to user experience and find that review solicitation increases volume but reduces average helpfulness, while refund decisions serve as credibility signals that elevate helpfulness ratings. Taken together, these results highlight how both user experience and platform interventions influence the quality and credibility of user-generated content.

5 - Accuracy, Awareness, and Acceptance: How Attention Drives the use of Point Versus Interval Forecasting Algorithms

Rob Basten, Eindhoven University of Technology, Eindhoven, Netherlands, Bas van Oudenhoven, Philippe van de calseyde, Evangelia Demerouti

Many operations rely on forecasts of future events, with the allocation of substantial amounts of money depending on these forecasts' accuracy. Fortunately, algorithmic forecasts often achieve high levels of accuracy. Still, many decision-makers reject these algorithms upon observing they are not perfectly accurate. What determines whether decision-makers perceive an algorithm as inaccurate, and how are these perceptions related to their willingness to use the algorithm when forecasting future events? We propose that the tendency to dismiss algorithmic advice is partially due to the common practice of presenting forecasts as single points rather than intervals. In Experiment 1, we find that people perceive equally imperfect forecasts as more accurate when these forecasts are presented as intervals rather than points. In Experiment 2, we examine how these differential accuracy perceptions influence people's willingness to use algorithms. Here, we find that people do not necessarily favor interval over point forecasts when deciding whether to rely on the algorithm to make predictions. Instead, such a preference only occurs when people are explicitly prompted to evaluate the algorithm's accuracy before deciding whether to commit to its predictions or not. Together, these findings underscore the complex processes underlying people's perceptions of algorithmic errors and their subsequent reliance on algorithms. Moreover, they show how managers can implement interval forecasts to effectively increase human forecasters' reliance on forecasting algorithms..

SD65

Building B Level 4 B405

Economics of AI

Invited Session

Artificial Intelligence

Chair: Stephanie Lee, University of Washington, Seattle, WA, United States

Co-Chair: Zibo Liu, Fudan University, Shanghai, 200433

1 - AI Pricing, Algorithmic Collusion, and Topological Structure

Yingying Chen, University of Washington, Seattle, WA, United States, XINGCHEN XU, Stephanie Lee, Yong Tan

Recent studies have extensively examined algorithmic collusion arising from AI-driven pricing algorithms, typically under the assumption that each seller competes directly with every other seller. However, in many practical contexts, competitive interactions are constrained by an underlying network topology. To address this, we develop a general repeated-game framework defined over arbitrary topological structures. Through extensive numerical experiments, we demonstrate how network topology shapes both dynamic pricing behavior and equilibrium outcomes, and we identify the key mechanisms driving these effects.

2 - Emotion in Echo: Understanding the Effectiveness of Group AI Emotional Support

Yuqing Hu, University of Washington, Seattle, WA, United States, Junyu Yao, Wendao Xue, Yifan Yu, Yong Tan

The use of artificial intelligence (AI) in emotional support conversations has received significant attention in both research and industry. While most studies have examined interactions with a single AI agent, little is known about the effects of interacting with a group of AI agents—a form of support that is beginning to emerge in practice. Drawing on social network and social support literature, we propose that merely increasing the number of AI agents may lead to inferior emotional support. Moreover, effective coordination among these agents in delivering complementary social support plays a critical role. To test these proposals, we conduct experiments in which participants interact with different AI systems designed to provide emotional support. This study aims to identify the potential of group AI in emotional support and contribute to both AI-in-emotional-support research and the broader body of social support literature.

3 - AI in Job Interviews: Promise OR Peril?

Zhen Fang, Fudan University, Shanghai, China, People's Republic of, Zibo Liu

Abstract: With the rapid development of artificial intelligence (AI), its application in job interviews has attracted firms seeking greater efficiency and lower hiring costs. While AI algorithms can effectively evaluate candidates with common characteristics, they often fall short when assessing applicants with niche or uncommon profiles due to lack of training data and the nature of AI algorithms. This study compares AI-led and human-led interview processes. By illustrating the potential weaknesses of AI in interview settings, we contribute to a more comprehensive understanding of its role in recruitment. Our study sheds light on the merits and drawbacks of AI in job interviews, and we provide managerial implications for firms aiming to integrate AI into their hiring practices.

4 - Search with AI-Simulated Future State Visuals: Evidence from the Pre-Owned Housing Market

Zhenbin Yan, Tongji University, Shanghai, China, People's Republic of, Shunyu Zhang

Customer search is challenging when future consumption experiences diverge significantly from a product's current state, such as in the pre-owned housing market where buyers often plan renovations. This makes search difficult, as customers must evaluate a property not just as it is but based on an uncertain vision of what it could become. While traditional mental simulations are often abstract and biased, advanced artificial intelligence (AI) technologies can simulate virtual product anticipation. This study investigates how AI-generated visualizations of renovated properties impact buyer search and market outcomes. Using data from a leading real estate platform, we find that AI virtual renovation reduces time-on-market by 9.3% and increases selling prices by 1.4%. However, there are underlying nuances. AI is most effective when information is balanced—neither too many options about the property's future nor too little detail about its current state. The effect is stronger for higher-value properties where anticipation is crucial for high-stakes choices and greater in markets with higher income and education, suggesting differential AI receptivity. In addition to online search, AI also increases online-to-offline conversions. This research highlights AI's potential to bridge the gap between current attributes and uncertain future experiences, thereby improving customer search and decision-making.

5 - More Human than Human: Empirical Evidence of the Effect of AI Generative Content in Online Communities

Jun Tan, University of Washington, Seattle, WA, United States, Lijia Ma, XINGCHEN XU, Yong Tan

AI generation tools have been widely used since their invention. However, how AI-generated content shapes discussions in online communities remains under-explored. In this paper, we conduct a field experiment to examine how an exogenous AI-generated content intervention affects subsequent discussion in a leading online forum.

6 - Rethinking Data Value: An Asymmetric Shapley Framework for Temporal and Context-Aware Valuation in Modern AI Systems

Xi Zheng, University of Washington, Seattle, WA, United States, Yinghui Huang, Xiangyu Chang, Ruoxi Jia, Yong Tan

As data emerges as a vital driver of technological and economic advancements, a key challenge is accurately quantifying its value in algorithmic decision-making. The Shapley value, a well-established concept from cooperative game theory, has been widely adopted to assess the contribution of individual data sources in supervised machine learning. However, its symmetry axiom assumes all players in the cooperative game are homogeneous, which overlooks the complex structures and dependencies present in real-world datasets. To address this limitation, we extend the traditional data Shapley framework to asymmetric data Shapley, making it flexible enough to incorporate inherent structures within the datasets for structure-aware data valuation. We also introduce an efficient k-nearest neighbor-based algorithm for its exact computation. We demonstrate the practical applicability of our framework across various machine learning tasks and data market contexts.

SD66

Building B Level 4 B406

Application and Cautionary Notes of GenAI usage in OR/Analytics Classroom

Invited Session

INFORMS Committee on Teaching and Learning

Chair: Moses Chan, Northwestern University, Evanston, IL, 60208, United States

1 - Using OR Losing: Student Experiences with AI Tools and Knowledge Acquisition in Data Visualization

Rich Ross, University of Virginia, Charlottesville, VA, United States

Generative AI tools have been thrust into the hands of undergraduate students. Some courses and instructors deal with these tools by forbidding their use entirely; others turn a blind eye. Some have begun to adopt their use in or out of the classroom. One big question on the minds of instructors and students is: "What is the consequence of using GenAI throughout a course?"

In this talk, I'll approach GenAI use in Statistics courses from three perspectives. First, I'll share insights into how students do (and do not) use AI tools in classes with varying policies. Second, I'll discuss how I've incorporated a pros and cons relationship with GenAI tools in my classes. This includes examples both of places where I encourage its use and and places where I illustrate its shortcomings. Lastly, I'll talk about lessons learned and provide concrete solutions that I've used and helped others to implement in an interdisciplinary setting, showing some lessons learned which are not topic-dependent (while recognizing that some are content-specific).

2 - Balancing Assistance and Integrity: How Students use Genai in Business Analytics

Huy Huynh, Emory University, Atlanta, GA, United States

As generative AI tools like ChatGPT become increasingly integrated into higher education, educators face the challenge of promoting student learning while maintaining academic integrity. This talk explores how undergraduate students in Business Analytics courses use generative AI to support their coursework, drawing on surveys, classroom observations, and assignment analyses. We examine the spectrum of student behaviors—from responsible use that enhances understanding to overreliance that risks undermining learning outcomes. By analyzing patterns of usage, we identify strategies students use to balance AI assistance with personal accountability. The session concludes with practical recommendations for instructors seeking to incorporate GenAI in a way that fosters ethical engagement, critical thinking, and skill development in analytics education.

3 - Introducing Undergraduate Business Students to Genai: An Update

Neil Desnoyers, Saint Joseph's University, Philadelphia, PA, United States

Starting Fall 2025 in the Haub School of Business at SJU, the required undergraduate “Intro to Information Systems” course has morphed into a required “AI in Business” course. Interim results and lessons learned will be discussed.

4 - Rejoinder: GenAI Usage and Precautions in OR and Analytics Classroom

Moses Chan, Northwestern University, Evanston, IL, United States

This talk reflects on the usage and precautions of generative AI in teaching operations research and analytics. This portion of the session focuses on discussion and exchanges of lessons learned in preparing for student success, balancing the reliance on GenAI, and promoting learning effectiveness.

SD67

Building B Level 4 B407

DM Best Student Paper Award

Award Session

Data Mining

Chair: Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

1 - Inv-Entropy: a Fully Probabilistic Framework for Uncertainty Quantification In Language Model

Haoyi Song, University of Michigan, Ann Arbor, Ann Arbor, MI, United States, Ruihan Ji, Naichen Shi, Fan Lai, Raed Al Kontar

We model LLM input-output pairs via dual random walks through perturbations. Based on this, we propose a probabilistic framework featuring an inverse model and introduce Inv-Entropy, a novel uncertainty measure. We also present TSU (Temperature Sensitivity of Uncertainty), a new evaluation metric that directly assesses uncertainty without relying on correctness.

2 - The Power of Adaptivity in Experimental Design

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

"Given n experiment subjects with potentially heterogeneous covariates and two possible treatments, namely active treatment and control, this paper addresses the question of determining the optimal accuracy in estimating the treatment effect. We propose an experimental design that approaches this optimal accuracy, giving a (non-asymptotic) answer to this fundamental question. The methodological contributions are as follows. First, we establish an idealized optimal estimator with minimal variance as benchmark, and then demonstrate that adaptive experiment is necessary to achieve near-optimal estimation accuracy. Second, by incorporating a doubly robust method into sequential experimental design, we frame the optimal estimation problem as an online bandit learning problem, bridging the fields of statistical estimation and bandit learning. Finally, using tools and ideas from both bandit algorithm design and adaptive statistical estimation, we propose a general low switching adaptive experiment framework, which could be a generic research paradigm for a wide range of adaptive experimental designs. Through novel lower bound techniques for non-i.i.d. data, we demonstrate the optimality of our proposed experiment. Numerical result indicates that the estimation accuracy approaches optimal with as few as two or three policy updates."

3 - Distributionally Robust Universal Classification: by Passing the Curse of Dimensionality

Siyuan Chen, Georgia Institute of Technology, Atlanta, GA, United States, Weijun Xie

The Distributionally Robust Universal Classification (DRUC) framework addresses overfitting in empirical risk minimization by incorporating Wasserstein-based robustness. We propose an in-sample DRUC reformulation, establish asymptotic and finite-sample guarantees, develop a MILP model, and design a 2-approximation algorithm. Numerical results confirm its efficiency and strong out-of-sample performance.

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

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

We study the fundamental problem of how many samples are needed to learn an ϵ -optimal policy in inventory control with setup costs. Our proposed algorithm achieves superior sample complexity than that achieved by either the plug-in demand estimator or directly applying the recently developed optimal reinforcement learning algorithms.

SD68

Building B Level 4 B408

Optimization and Surrogate Methods for Black-Box Systems

Invited Session

Data Mining

Chair: Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

1 - Lipschitz Based Lower Bound Construction for Surrogate Optimization

Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States, Mohammadsina Almasi, Jay Rosenberger

Bounds play a vital role in guiding optimization algorithms by enhancing convergence, improving solution quality, and quantifying optimality gaps. While Lipschitz-based lower bounds are well-established, their effectiveness is often constrained by the function's topological properties. To address these limitations, we propose an approach that integrates nonlinear distance metrics with surrogate approximations, yielding more adaptive and accurate lower bounds. A key aspect of our methodology lies in the flexibility of the chosen distance metric, which can be adapted to various function behaviors. In particular, we explore sublinear and superlinear metrics under the Hölder continuity assumption, demonstrating their capacity to capture local function characteristics beyond the scope of conventional linear bounds. Empirical evaluations on diverse benchmark test problems show that our approach surpasses both standard Lipschitz-based and statistical methods in producing high-quality lower bounds. We further employ these refined bounds as an acquisition function within surrogate optimization, a common technique for expensive black-box problems wherein a surrogate model approximates the true objective function. By leveraging these bounds to balance exploration and exploitation, we effectively prioritize evaluations in regions with high potential for improvement. Overall, our framework not only offers more accurate and flexible lower bound estimates but also acts as a robust acquisition strategy that expedites convergence to near-optimal solutions in black-box optimization.

2 - Fairness-Aware Bandit Optimization with Delayed Feedback

Mohammadsina Almasi, University of Illinois at Chicago, Chicago, IL, United States, Hadis Anahideh

Decision-making under uncertainty with limited resources poses critical challenges in domains such as education, healthcare, and social services, particularly when the outcomes of actions are not observed immediately but manifest with delay, shaped by hidden dynamics beyond observed features. These settings introduce complex constraints: limited budgets, equity considerations, and the need to model long-term impact. Traditional approaches that assume immediate feedback and static environments are insufficient in capturing the nuanced, temporal structure of such problems. We propose a novel decision-making framework based on delayed-feedback multi-armed bandits that integrates a neural network trained on individual-level observational data. Rather than modeling reward at the level of individual actions alone, the framework learns latent factors that influence delayed outcomes and aggregates these signals to infer subgroup-level reward distributions. By modeling how past allocation patterns influence future responses, the algorithm learns a dynamic, temporally-aware policy that evolves in response to prior actions. The learned policy satisfies both budgetary constraints and fairness criteria by ensuring balanced performance across subgroups while optimizing overall outcomes. Our method provides a data-driven, adaptive mechanism for learning effective allocation strategies under uncertainty and delay, addressing both the structural and ethical demands of real-world policy problems. Empirical results on a large-scale educational dataset demonstrate the framework's ability to capture subgroup-specific reward dynamics, model temporal dependencies in delayed feedback, and achieve superior trade-offs between outcome quality and fairness in constrained environments.

3 - SO0GO: a Surrogate-Based Zero-Th Order Global Optimization Toolbox for Expensive Black-Box Problems

Juliane Mueller, National Renewable Energy Laboratory, Golden, CO, United States, Wesley da Silva Pereira

We present a toolbox for derivative free global optimization aimed at computationally expensive black-box problems. The toolbox integrates several of our previously developed derivative free optimization algorithms including continuous, mixed-integer, and multi-objective developments. The methods are based on the use of surrogate models (radial basis functions and Gaussian process models) that allow us to selectively query the true expensive objective functions only at the most promising points in the search space, thus significantly reducing the computational expense of the optimization. Several acquisition functions are implemented including approaches for batch sampling. The toolboxes is freely available at <https://github.com/NREL/soogo/tree/main>.

4 - Enhancing Bayesian Optimization with Large Language Models for Black-Box Problems

Andrea Carbonati, University of Illinois Chicago, Chicago, IL, United States, Hadis Anahideh

Bayesian Optimization (BO) is widely recognized for its effectiveness in optimizing expensive, black-box functions, particularly in experimental and computationally intensive contexts. However, traditional BO faces challenges in efficiently navigating complex, high-dimensional search spaces and leveraging rich domain-specific heuristics. Large Language Models (LLMs) have recently emerged as powerful heuristic reasoners capable of deriving context-rich insights through natural language interactions. Nevertheless, standalone LLM-driven optimization suffers from numerical inaccuracies, instability due to prompt sensitivity, and difficulty scaling to purely numeric optimization tasks. To overcome these limitations, we propose a robust hybrid framework integrating the heuristic reasoning strengths of LLMs with the rigorous statistical modeling capabilities of BO. Within this framework, LLMs proactively generate contextually informed priors and candidate solutions, effectively guiding the BO surrogate model towards promising regions in the solution space, while BO ensures robust numeric precision and rigorous uncertainty quantification. Empirical validation across diverse real-world benchmarks—including hyperparameter tuning and costly simulation-based optimization tasks—demonstrates substantial improvements in convergence speed, sample efficiency, and solution quality compared to traditional BO methods. This research paves the way for more effective, scalable, and practically applicable BO solutions, harnessing the complementary strengths of statistical rigor and contextual heuristic intelligence.

SD69

Building B Level 4 B409

Behavioral Patterns and Choices in Innovation

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Tian Chan, Emory University, Atlanta, GA, United States

1 - AI-Based Group: The Impact of AI Advice on Group Decision-Making Process

Qiong Xia, INSEAD, FONTAINEBLEAU, France, Vivek Choudhary, Yash Shrestha, Sezer Ulku

Group decision-making often involves a discussion process where individuals exchange ideas, discuss issues, and influence each other's opinions. With the growing adoption of AI in organizations, understanding its impact on this process is critical. Using a laboratory experiment where group members form their own predictions and engage in discussions, we compare the effects of AI advice introduced individually before (Early AI) and after (Late AI) discussion, as well as AI advice introduced collectively to the group before discussion (Group AI), to scenarios where there is no AI advice (No AI).

The results reveal that groups receiving AI advice outperform No AI groups, with Early AI achieving the highest prediction accuracy. However, group dynamics vary depending on whether AI is introduced before or after group discussions, regardless of whether it is introduced individually or collectively. When AI advice is introduced before group discussion, groups in both Early AI (individual) and Group AI (collective) engage in shorter discussions to reconcile the differences in individuals' predictions compared to No AI groups. Additionally, these groups discuss less the features of the prediction variables, leading to increased prediction errors in their own predictions over time compared to No AI groups. In contrast, when AI is introduced after group discussion (Late AI), it disrupts group dynamics by increasing disagreement on final predictions, thereby reducing consensus compared to No AI groups. However, Late AI groups exhibit significantly greater cognitive engagement in performance feedback.

2 - Team Building and Incentive Schemes in Collaborative Projects

Ignacio Rios, The University of Texas at Dallas, Richardson, TX, United States, Ruth Beer, Anyan Qi

Worker-induced delays pose a major challenge in collaborative project environments. This paper investigates how incentive structures and managerial interventions—particularly team-building activities—can reduce such delays. We focus on two prevalent incentive schemes: (i) individual incentives, which reward workers based on their own performance, and (ii) group incentives, which tie rewards to overall team outcomes. We hypothesize that the effectiveness of group incentives hinges on the strength of interpersonal ties within teams and explore how team-building activities may enhance their impact.

We develop a game-theoretic model of a two-worker project involving sequential tasks, where each worker decides whether to proceed with their task or delay it by pursuing an alternative option. The model predicts that delays are more likely when workers face attractive outside options or when there is excess slack in the project timeline. Group incentives can reduce delays, but only when workers care about each other's payoffs. We then test these predictions in a laboratory experiment, varying both the incentive scheme and the presence of a team-building activity.

The results confirm the behavioral regularity of *slack-based procrastination*: workers are more likely to delay tasks when more slack is available. Importantly, we find that team-building activities reduce delays by nearly 10% under group incentives—but not under individual incentives—highlighting a strong interaction effect. These findings suggest that pairing group incentives with light-touch interventions like team-building can meaningfully reduce delays in collaborative settings, and that the timing and structure of such interventions are critical for their success.

3 - What Brings Lurkers to the Online Table: The Role of Goal Gradient Effect

Shizhen Chen, Emory University, Atlanta, GA, United States, Tian Chan, Anandhi Bharadwaj

Online communities (OCs) are dominated by lurkers—individuals who consume content without contributing—yet collective action projects, which harness dispersed contributions to generate public goods, drive innovation, and foster civic engagement, depend critically on engaging these silent participants. Understanding how to activate lurkers is therefore vital to the success of citizen science initiatives, crowdfunding campaigns, online petitions, and similar efforts. Drawing on data from Reddit's r/place 2022 social experiment and goal-gradient theory, we argue that clearly defined, singular, and achievable objectives function as powerful visual cues that prompt latent users to take action.

We test three main hypotheses in this setting. First, greater goal completion accelerates a project's productive work rate (H1a), an effect that is stronger when goals are presented with higher visual complexity (H1b). Second, this goal-completion effect is most pronounced in communities with large latent user bases, suggesting that realistic goals primarily spur lurker engagement rather than merely boosting activity among existing contributors (H2). Third, significant setbacks diminish the goal-gradient effect, indicating the importance of dynamically recalibrating objectives to maintain momentum (H3).

Our results offer actionable guidance for OC managers: maintain high goal visibility, adapt targets in response to setbacks, and leverage visual complexity to mobilize the silent majority essential to collective online endeavors.

4 - Cultures for Innovation

Rob Glew, McGill University (Desautels), Montréal, QC, Canada, Jeremy Hutchison-Krupat

Many organizations face the challenge of pursuing novel innovation, where the actions required to uncover the technology are unknown. As management of innovation projects is typically delegated to an R&D unit, ensuring resources are committed to experiments with the greatest chance of success is a central concern for senior management. Yet, their ability to do so is limited, as the specific steps to take cannot be articulated and the productivity of experiments is only known ex-post, creating a risk of moral hazard. In this paper, we examine how an organization's innovation culture may be used to resolve this problem. We study the situation where senior management delegates operational decisions for innovation to an R&D manager, whose efforts and actions cannot be fully verified. Instead, senior management may only set targets for the resources allocated to derivative innovation (which can be verified) and their tolerance for failure, creating the organizations' innovation culture. Recognizing that, in practice, organizations vary in terms of the autonomy and tolerance they assign to innovation projects, we analyze both cultural levers separately and in combination. Our results provide an alternative view of tolerance for failure by including moral hazard: reducing tolerance for failure can enable organizations to pursue riskier innovation projects. We verify that setting targets for derivative innovation can increase the rate of novel innovation, but only for a subset of projects. Otherwise the cost of such measurement may be unjustified. Moreover, some organizations may benefit from prohibiting derivative innovation to pursue riskier projects.

Sunday, October 26, 4:15 PM - 5:30 PM

SD36

Building B Level 2 B210

Advances in Nonlinear Optimization II

Invited Session

OPT: Nonlinear Optimization

Chair: Raghu Bollapragada, The University of Texas at Austin, Austin, TX, United States

1 - Ransac Revisited: An Improved Algorithm for Robust Subspace Recovery Under Adversarial and Noisy Corruptions

Guixian Chen, University of Michigan, Ann Arbor, MI, United States, Jianhao Ma, Salar Fattahi

In this paper, we study the problem of robust subspace recovery (RSR) in the presence of both strong adversarial corruptions and Gaussian noise. Specifically, given a limited number of noisy samples---some of which are tampered by an adaptive and strong adversary---we aim to recover a low-dimensional subspace that approximately contains a significant fraction of the uncorrupted samples, up to an error that scales with the Gaussian noise. Existing approaches to this problem often suffer from high computational costs or rely on restrictive distributional assumptions, limiting their applicability in truly adversarial settings. To address these challenges, we revisit the classical random sample consensus (RANSAC) algorithm, which offers strong robustness to adversarial outliers, but sacrifices efficiency and robustness against Gaussian noise and model misspecification in the process. We propose a two-stage algorithm, RANSAC+, that precisely pinpoints and remedies the failure modes of standard RANSAC. Our method is provably robust to both Gaussian and adversarial corruptions, achieves near-optimal sample complexity without requiring prior knowledge of the subspace dimension, and is more efficient than existing RANSAC-type methods.

2 - Fair Supervised Learning via Smooth Nonconvex Constraints

Zahra Khatti, Lehigh University, Bethlehem, PA, United States, Daniel Robinson, Frank E. Curtis

We propose a new strategy for fair supervised machine learning, emphasizing computational efficiency and reliable fairness enforcement. Our approach introduces a novel smooth nonconvex surrogate to approximate discontinuous unfairness measures, ensuring tight and effective fairness guarantees compared to conventional convex approximations. Unlike traditional regularization methods, our strategy employs hard constraints, allowing explicit control over fairness tolerances without extensive tuning of regularization parameters. Additionally, our framework efficiently handles multiple, potentially conflicting fairness criteria simultaneously, resulting in tractable optimization problems with minimal parameter tuning.

3 - A Bilevel Optimization Method Using Synthetic Data to Mitigate Unfairness and Preserve Privacy in Collaborative Machine Learning

Chia-Yuan Wu, Lehigh University, Bethlehem, PA, United States, Frank E. Curtis, Daniel Robinson

We propose a two-stage strategy for collaborative machine learning that promotes fair predictions, prevents client-data leakage, and reduces communication costs without iterative communication. In the first stage, each client generates a synthetic dataset by solving a bilevel optimization problem aimed at ensuring that the global model yields fair predictions. In the second stage, a differentially private method is applied to the first-stage synthetic data to obtain a second synthetic dataset. Each client then sends its second-stage synthetic data to the server. These data are used to train the global model using conventional machine learning techniques, eliminating the need to handle fairness-specific aggregation while preserving privacy. Our approach requires only a single communication round, is communication-efficient, maintains data privacy, and promotes fairness. Empirical results demonstrate that our method effectively reduces unfairness while preserving client privacy through synthetic data.

4 - Higher Order Polynomial Model-Based Derivative-Free Methods

Abraar Chaudhry, Georgia Institute of Technology, Atlanta, GA, United States, Katya Scheinberg

Traditional model-based derivative free methods, such as those pioneered by Powell, iteratively construct second-order interpolation models of the objective function and optimize these models over trust-regions. Higher order polynomials may provide better models, however, they are hard to optimize. We propose a new approach of constructing and optimizing higher degree polynomial models for derivative free optimization. This is done using techniques from polynomial optimization and sum of squares. We will discuss the practical and theoretical properties of our method.

5 - Zeroth Order Schemes for Nonsmooth, Nonconvex, Stochastic Optimization

Luke Marrinan, Penn State University, State College, PA, United States, Farzad Yousefian, Uday Shanbhag

We consider the minimization of a Lipschitz continuous and expectation-valued function over a closed and convex set. Our focus lies on obtaining both asymptotics as well as rate and complexity guarantees for computing an approximate stationary point (in a Clarke sense) via zeroth-order schemes. We adopt a randomized-smoothing-based approach reliant on minimizing a smooth approximation of our objective function. In such a setting, we develop a zeroth-order stochastic gradient and quasi-Newton schemes reliant on a combination of randomized and Moreau smoothing is proposed and analyzed, for which iteration and sample complexities are derived.

SE01

Building A Level 3 A301

George Nicholson Student Paper Competition

Award Session

Nicholson Student Paper Competition

Chair: Xinyun Chen, Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A

Co-Chair: Wolfram Wiesemann, Imperial College Business School, London, United Kingdom

1 - Multi-Agent Markov Entanglement

Shuze Chen, Columbia University, New York City, NY, United States

Value decomposition is fundamental in multi-agent reinforcement learning, yet its theoretical justifications remain uncharted. This work establishes that a Markov decision process (MDP) permits exact value decomposition iff. it's not 'entangled'—a concept akin to quantum entanglement. We further introduce the measure of Markov

entanglement, quantifying decomposition error in general MDPs. Using this framework, we prove a widely-used class of index policies is weakly-entangled. Finally, we show efficient estimation of Markov entanglement, guiding practitioners on the decomposition feasibility.

2 - A Simple Uniformly Optimal Method Without Line Search for Convex Optimization

Tianjiao Li, Georgia Institute of Technology, Atlanta, GA, United States

In this paper, we present a novel auto-conditioned fast gradient method (AC-FGM) that can achieve an optimal $O(1/k^2)$ rate of convergence for smooth convex optimization without requiring the estimate of a global Lipschitz constant or the employment of line search procedures. We then extend AC-FGM to solve convex optimization problems with Hölder continuous gradients and show that it automatically achieves the optimal rates of convergence uniformly for all problem classes with the desired accuracy of the solution as the only input. Finally, we report some encouraging numerical results.

3 - Dynamic Matching with Post Allocation Service and Its Application to Refugee Resettlement

Soonbong Lee, Yale University, New Haven, CT, United States

Motivated by our collaboration with a major U.S. refugee resettlement agency, we study a dynamic matching problem where each arrival must be matched immediately and irrevocably to a static resource and require post-allocation services from a dynamic resource. The central planner aims to balance matching rewards with the costs of congestion and over-allocation. We design learning-based algorithms that are distribution-free, asymptotically optimal in certain regimes, and computationally efficient. When tested on real data from our partner agency, our methods outperform the existing ones.

4 - Maximum Load Assortment Optimization: Approximation Algorithms and Adaptivity Gaps

Marouane Ibn Brahim, Cornell Tech / Cornell University, New York, NY, United States

We study Maximum Load Assortment Optimization, where the goal is to steer a stream of customers toward selecting the same product from an offered assortment. Motivated by applications like Attended Home Delivery and group scheduling, each customer chooses a product or opts out, with choices governed by the Multinomial Logit model. The decision maker selects assortments to maximize the expected maximum number of selections for any single product. We present near-optimal algorithms for both static and dynamic versions of the problem, addressing key computational challenges.

SE02

Building A Level 3 A302

Thriving in Industry: Leveraging AI and Analytics for Career Success

Panel Session

Committee's Choice

Co-Chair: Shixiang Woody Zhu, Carnegie Mellon University, Pittsburgh, PA, United States

Co-Chair: Jia Liu, University of Florida, Gainesville, FL, United States

1 - Moderator Panelist

Jia Liu, Auburn University, Auburn, AL, United States

This panel session is part of the committee's choice sessions of INFORMS. The session, co-organized Data Mining (DMS) and Quality, Statistics, and Reliability (QSR) Societies, will bring together leaders from the industry to share their journeys, lessons learned and insights on how AI and analytics can shape and accelerate impactful careers. The session welcomes a broad audience of academics, practitioners, and students.

2 - Moderator Panelist

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

3 - Panelist

Yigit Arisoy, Amazon, Marietta, GA, United States

4 - Panelist

Shan Ba, LinkedIn, Mountain View, CA, United States

5 - Panelist

Ryan Lekivetz, JMP Statistical Discovery LLC, Cary, NC, United States

6 - Panelist

Liu Richard, Wells Fargo, Charlotte, NC, United States

7 - Panelist

Xin Ma, Monks, Atlanta, GA, United States

SE03

Building A Level 3 A303

Building MCDM/A models: Practical and Methodological Issues

Invited Session

Multi Criteria Decision Making

Chair: Eduarda Frej, Universidade Federal de Pernambuco, Recife, Brazil

Co-Chair: Danielle Costa Morais, Universidade Federal de Pernambuco, Recife, N/A, Brazil

1 - Analyzing the Effects of Holistic Evaluation in the Fitradeoff Method: Results of A Simulation Study

Eduarda Frej, Universidade Federal de Pernambuco, Recife, Brazil, Manoel Ribeiro, Tarsila Vasconcelos, Adiel De Almeida

The FITradeoff method combines two types of preference modelling in its structure: elicitation by decomposition and holistic evaluation. This combination enhances the flexibility of the method, improving the preferences elicitation task in a sense that fewer elicitation questions are expected to be answered by the Decision Maker (DM) to find a recommendation for the multicriteria decision problem. In this context, this work aims to evaluate the impact of conducting holistic evaluations in FITradeoff, evaluating the potential for reduction in the number of questions answered by the DM to achieve a recommendation. We present results of a simulation study, carried out considering different scenarios for number of criteria, number of alternatives and weights distribution pattern. Hypotheses on the reduction of the number of questions were tested for identifying the actual effect of performing HE after the step of the ranking of criteria weights, in the beginning of the elicitation process. The results have shown that holistic judgments have a significant impact on the reduction of the number of questions answered by the DM in the elicitation process, in a sense that the total effort spent by DMs reduces in a significant manner, improving the decision process.

2 - An Analysis of the Heuristics Associated with the Decomposition Elicitation Procedure of the Fitradeoff Method

Jonatas Almeida, Federal University of Pernambuco, Caruaru, Brazil, Eduarda Frej, Adiel de almeida

The FITradeoff method, which belongs to the MAVT set of methods, combines two approaches to flexible elicitation of scale constants, the decomposition procedure and the holistic evaluation procedure. The first approach, decomposition, is based on the tradeoff procedure, which has a very robust axiomatic structure, but which can

prolong elicitation and make it cognitively challenging, leading to mistakes in the decision maker's statements. The flexible approach of the FITradeoff decomposition procedure allows the use of partial information to evaluate alternatives according to the information needed for the problem, making elicitation faster and easier. However, elicitation efficiency is still an important objective, so methodological improvements that seek to speed up elicitation and provide greater flexibility to the decision maker are important contributions. In this context, an adjustment at the beginning of the decomposition procedure allows the use of a heuristic that speeds up elicitation, with the potential to significantly reduce the number of questions asked to the decision maker. The heuristic acts by accelerating the reduction of the feasible region of the weight space; however, its efficiency depends on the preference scenario.

3 - Artificial Intelligence in Multi-Criteria Decision Making/Aid Methods: A Systematic Literature Review

BRUNO CICCUI, Federal University of Pernambuco, Recife, Brazil, Eduarda Frej, Adiel T. de Almeida

Over the past two decades, Artificial Intelligence (AI) has experienced substantial development, progressing from traditional predictive models and human-like computational methods to becoming an integral part of complex decision-making processes. A notable contemporary trend is the integration of AI with Multi-Criteria Decision-Making/Aid (MCDM/A) methods, aimed at providing more effective and intelligent decision support in multifaceted and uncertain environments. This study presents a systematic literature review that explores how AI techniques have been integrated with MCDM/A methods. Based on 106 peer-reviewed articles sourced from major scientific databases, the review is structured around six research questions and employs both bibliometric analysis and qualitative content assessment to uncover prevailing patterns, applications, and gaps in the literature. The findings highlight several benefits of combining AI with MCDM/A methods, including the ability to handle imprecise and vague data, manage high-dimensional information, and conduct predictive modeling and scenario-based simulations. These capabilities are particularly useful in engineering fields such as logistics and intelligent transportation systems, production and resource allocation, energy systems, and environmental planning. Despite these benefits, the review identifies critical challenges related to data quality and resolution, explainability and biases of learning algorithms, complexity and scalability of AI algorithms, and concerns related to human-AI interactions within the AI-enhanced MCDM/A methods. In summary, this paper offers a comprehensive overview of the synergistic relationship between AI and MCDM/A methods, providing insights into how this integration can overcome existing limitations and support robust, efficient decision-making in engineering and other complex domains.

4 - Public Procurement and Social Sustainability: a Multicriteria Decision-Making Model for Contractor Selection

George da Mota Passos Neto, UFPE, Recife, Brazil, Luciana Alencar, Rodolfo Valdes-Vasquez

Traditional contractor selection practices in public procurement often overlook broader societal impacts, despite their critical role in ensuring successful and sustainable project outcomes. This study proposes a model for public procurement that integrates social sustainability, improving transparency and alignment with societal values of contractor selection. Social sustainability criteria were identified through a literature review and include stakeholder participation, occupational safety and health, employment, community impacts, socioeconomic compliance, environmental impact, neighborhood amenities, user considerations, diversity, cultural heritage, supply chain, ethics, and innovation. The model integrates Multi-Criteria Decision-Making/Aiding (MCDM/A) methods and emphasizes the importance of the decision-maker (DM) in this process. The FITradeoff method, an MCDM/A approach based on partial information, was employed to reduce the DM's cognitive effort and, consequently, lower the model's inconsistency rate. The primary contributions of the proposed model are twofold: i) incorporation of social sustainability criteria in public procurement for contractor selection, thus promoting accountability and societal well-being, and ii) offering practical techniques for managing decision-making complexities through partial information, thereby ensuring the contractor selection process is transparent, and aligned with societal goals. The study also presents an application of the model for selecting a contractor to design and construct a school in Brazil.

5 - Data-Driven Team Formation: a Machine Learning Approach for Effective Student Team Formation

Esmat Ghasemi Saghand, university of south florida, Tampa, FL, United States, Susana Lai-Yuen

Forming effective student teams significantly impacts collaborative learning and project success. However, traditional team formation methods rely heavily on predefined criteria or skills, which may not accurately capture the complex factors influencing actual team performance. These methods overlook valuable insights available from real-world team data, resulting in suboptimal team compositions and reduced effectiveness.

In this project, we propose a data-driven approach that leverages historical student survey data to build predictive machine learning models. The student data was obtained from CATME, a widely used team formation tool. The dataset includes individual student attributes such as GPA, weekly availability, English proficiency, and writing skills, combined with real-world team performance outcomes like final project scores.

We start by processing and converting qualitative CATME survey responses into structured numerical features. These individual attributes are then aggregated into comprehensive team profiles. Supervised learning models, including Random Forest and XGBoost, are trained to identify patterns predicting team success. Cross-validation techniques ensure the models' reliability, and methods such as SHAP analysis reveal which attributes most strongly correlate with effective team performance. The most influential features identified through this analysis can be explicitly defined and incorporated into future team formation processes.

Using the trained model, we identify key student attributes that positively impact team performance based on final project outcomes. Our analysis reveals how these factors correlate with successful collaborations, enabling instructors to prioritize attributes most predictive of effective teamwork. By leveraging these insights, student teams can be formed that improve performance outcomes and foster collaborative interactions.

6 - From Leaf to Logistics: Multi-Objective Optimization and Scenario Analysis in Tea Supply Chains

Zehra Melis Teksan, ICRON , Amsterdam, Netherlands

Managing blending and procurement operations in the food industry involves complex trade-offs among cost, quality, and customer satisfaction. This presentation details the real-world implementation of an advanced decision support system for a leading tea producer. The solution leverages multi-objective optimization, machine learning, and real-time scenario analysis within a collaborative, multi-user environment. Designed to support both medium- and long-term planning, the system enables accurate operational cost forecasting and dynamic decision-making. By integrating these advanced analytics capabilities, the platform significantly enhances the agility and profitability of blending and procurement processes, while boosting customer satisfaction. The case illustrates how such a system can deliver substantial strategic value to companies in the blended food product sector.

SE04

Building A Level 3 A304

Financial Engineering and FinTech

Invited Session

Finance

Chair: Xuedong He, The Chinese University of Hong Kong, Shatin, Hong Kong

1 - Asset Pricing with Higher-Order Dependence in Labor and Stock Markets

Wei Jiang, Hong Kong University of Science and Technology, Hong Kong, Hong Kong

This study investigates higher-order dependence, which captures the joint occurrence of extreme values in labor and stock markets, and its implications for asset pricing. Higher-order dependence significantly affects household portfolio decisions by altering wealth risk profiles through skewness and kurtosis channels. Incorporating this dependence into asset pricing frameworks reveals that households reduce risky asset holdings and adjust participation thresholds, while ignoring it leads to economic inefficiencies and increased wealth inequality. These findings emphasize the critical role of higher-order dependence in shaping asset pricing models, household financial behavior, and the broader dynamics of labor and stock markets.

2 - Adaptive Window Selection for Financial Risk Forecasting

Ruodu Wang, University of Waterloo, Waterloo, ON, Canada

Risk forecasts in financial regulation and internal management are calculated through historical data. The unknown non-stationarity of financial data poses a substantial challenge in selecting an appropriate look-back window for risk modeling and forecasting. We develop a statistical learning method based on a stability principle to adaptively determine a window size for risk forecasting. A central component of the proposed methodology is to compare the realized scores with a threshold, which we evaluate based on an idea of bootstrap. The proposed method is applicable to the forecast of risk measures that are elicitable individually or jointly, such as Value-at-Risk (VaR) and the pair of VaR and Expected Shortfall (ES). Through simulation studies and empirical analyses, we demonstrate that the proposed approach outperforms the standard rolling window approach and a recently developed online learning approach, especially when there are structural changes in the data-generating process.

3 - Arbitraging on Decentralized Exchanges

Yutian Zhou, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Xuedong He, Chen Yang

Decentralized exchanges (DEXs) are alternative venues to centralized exchanges to trade cryptocurrencies (CEXs) and have become increasingly popular. An arbitrage opportunity arises when the exchange rate of two cryptocurrencies in a DEX differs from that in a CEX. Arbitrageurs can then trade on the DEX and CEX to make a profit. Trading on the DEX incurs a gas fee, which determines the priority of the trade being executed. We study a gas-fee competition game between two arbitrageurs who maximize their expected profit from trading. We derive the unique symmetric mixed Nash equilibrium and find that (i) the arbitrageurs may choose not to trade when the arbitrage opportunity and liquidity is small; (ii) the probability of the arbitrageurs choosing a higher gas fee is lower; (iii) the arbitrageurs pay a higher gas fee and trade more when the arbitrage opportunity becomes larger and when liquidity becomes higher; (iv) the arbitrageurs' expected profit could increase with arbitrage opportunity and liquidity. The above findings are consistent with our empirical study.

4 - Equity Crowdfunding vs. Ipos: Leveraging Dynamic Pricing for Blockchain Venture Success

Bretislav Hajek, National University of Singapore, Singapore, Singapore, Jussi Keppo, Steven Kou

Equity crowdfunding has emerged as a compelling alternative to traditional initial public offerings (IPOs) for blockchain-based ventures seeking early-stage capital. By issuing governance tokens—which may confer voting rights and participation in project direction—blockchain companies can bypass costly intermediaries such as investment banks, significantly reducing transaction costs associated with traditional IPO processes. Furthermore, blockchain-enabled equity crowdfunding empowers entrepreneurs to dynamically adjust pricing and marketing strategies in real-time, effectively responding to investor demand signals. This study analyzes the strategic advantages of equity crowdfunding relative to IPOs, emphasizing key benefits for both entrepreneurs and investors, while also considering challenges such as regulatory uncertainty and the evolving legal status of governance tokens as equity instruments.

5 - Dynamic Asset Allocation Under Monotone Additive Statistics in the Heston Model

Xuedong He, The Chinese University of Hong Kong, Shatin, Hong Kong, ZHAOLI JIANG, jianming xia

The monotone additive statistic is appealing due to its axiomatic foundation and its ability to exhibit behavior that is neither risk-averse nor risk-seeking. In this study, we apply the monotone additive statistic to dynamic portfolio selection within Heston's stochastic volatility model. To address the issue of time inconsistency induced by the monotone additive statistic, we consider equilibrium portfolio strategies that can be consistently implemented by the investor. We derive the equilibrium portfolio strategies for two scenarios, where preferences are characterized by the monotone additive statistics of terminal wealth and log return of investment, respectively. In both cases, we establish comparative results between the equilibrium strategy and the expected risk aversion strategy. Finally, our numerical studies reveal that the equilibrium strategy outperforms the expected risk aversion strategy when evaluated using monotone additive statistics.

SE05

Building A Level 3 A305

NSF Program Director Panel

Panel Session

NSF

Co-Chair: Tapas Das, University of South Florida, Tampa, FL, United States

1 - Panelist

Tapas Das, University of South Florida, Tampa, FL, United States

SE06

Building A Level 3 A306

Undergraduate Operations Research Prize

Award Session

Undergraduate Operations Research Prize

Chair: Ozge Yapar, Indiana University, Bloomington, IN, United States

1 - Reproducing Kernel Hilbert Space Choice Model

Yiqi Yang, Shanghai Jiao Tong University & The University of Texas at Austin, Shanghai, China, People's Republic of

We propose a flexible discrete choice modeling framework using vector-valued Reproducing Kernel Hilbert Spaces over choice-set and feature pairs, capturing context-dependent interactions without strong structural assumptions. We design kernels for both featureless and feature-based settings, enabling convex estimation via the Representer Theorem. For scalability, we incorporate random feature approximations and a Neural Tangent Kernel with multi-head attention. Our method achieves linear complexity and provable generalization, outperforming benchmarks on synthetic and real-world datasets.

2 - Exact Rounding for Fair K-Median Under Quasi-Convex Costs

Alireza Rezaeimoghadam, Sharif University of Technology, Tehran, Iran, Alireza Danaei

We study a constrained variant of the k-median problem imposing a partition matroid constraint to promote fairness. To make this NP-complete problem tractable, we assume that for each client, service costs are quasi-convex over an ordering of facilities. We design an exact poly-time rounding algorithm using a novel coloring scheme to convert any optimal fractional LP solution into an integral one. Our framework extends to settings with facility opening costs and robust multi-assignment variants. Experiments on healthcare facility placement in Alabama demonstrate strong empirical performance.

3 - Power Constrained Nonstationary Bandits with Habituation and Recovery Dynamics

Fengxu Li, University of Wisconsin–Madison, Madison, WI, United States

While existing algorithms minimize regret for personalized interventions, they often lack statistical power to infer population-level effects due to limited sampling. We introduce ROGUE-TS, a Thompson Sampling algorithm for ROGUE (Reducing or Gaining Unknown Efficacy) bandits which models non-stationary dynamics influenced by behavior effect such as habituation and recovery, along with a probability clipping method that balances personalization and power. Theoretical guarantees and validation on physical activity promotion data show improved regret and robust detection of treatment effects.

4 - Llm Serving with Variable Prefill and Decode Lengths

Meixuan Wang, Tsinghua University, Beijing, China, People's Republic of

We study scheduling LLM requests with heterogeneous input prompt and output token lengths to minimize total latency. We prove that the problem is NP-hard due to batching, memory constraints, and dynamic KV cache usage. We analyze FCFS and Shortest-First, showing that their competitive ratios are unbounded when the problem

scales up. To address this, we propose a novel algorithm, Sorted-F, with a provable competitive ratio of 48. We also develop practical heuristics (dynamic programming, local search, LP-based) that outperform baselines in simulations while remaining computationally efficient.

SE07

Building A Level 3 A307

Human Transportation Networks

Job Market Showcase

Job Market Showcase

Chair: Tianming Liu, University of Michigan, Ann Arbor, United States

1 - Learning to Generate Synthetic Human Mobility Data: a Physics-Regularized Gaussian Process Approach Based on Multiple Kernel Learning

Ekin Ugurel, University of Washington, Seattle, WA, United States, Shuai Huang, Cynthia Chen

Passively-generated mobile data has grown increasingly popular in the travel behavior literature. A relatively untapped potential for passively-generated mobile data is synthetic population generation, which is the basis for any large-scale simulations for purposes ranging from state monitoring, policy evaluation, and digital twins. And yet, this significant potential may be hindered by the growing sparsity or rate of missingness in the data, which stems from heightened privacy concerns among both data vendors and consumers (users of service platforms generating individual mobile data). To both fulfill the great potential and to address sparsity in the data, there is a need to develop a flexible and scalable model that can capture individual heterogeneity and adapt to changes in mobility patterns. We propose a conditional-generative Gaussian process framework that learns kernel structures characterizing individual mobile data and can provably replicate observed patterns. Our approach integrates physical knowledge to regularize the framework such that the generated data obeys constraints imposed by the built and natural environments (such as those on velocity and bearing). To capture travel behavior heterogeneity at the individual level, we propose a data-driven multiple kernel learning approach to determine the optimal composite kernel for every user. Our experiments demonstrate that: (1) the impact of kernel choice on mobility metrics derived from synthetic data is non-negligible; (2) physics-regularization not only reduces model bias but also improves uncertainty estimates associated with the predicted locations; and (3) the proposed method is robust and generalizes well to varying individuals and modes of travel.

2 - Choice-Based Transportation Network Design Using Hybrid Reweighted L1-Norm Minimization

Pushpendra Singh, Dartmouth College, Hanover, NH, United States, Vikrant Vaze

Network design is a well-known problem in transportation systems. Most of the existing transportation network design studies treat customer demand as exogenous. In reality, demand is impacted by the level of service, which is itself determined by the network design. A network that carries the passengers/goods from origin to destination with a lower travel time is expected to attract a higher demand. Thus, optimal network design affects and is affected by demand. Capturing this two-way relationship requires modeling endogenous customer choice, significantly increasing the problem complexity. We formulate a mixed-integer non-convex optimization model for network design that embeds a discrete choice customer demand model within it. Integrality arises from arc and path selection variables which are binary, while non-convexity stems from embedded choice models such as the logit model. We propose a general and compact formulation to represent this class of problems.

We propose the Reweighted Iterative Partial Penalty L1-norm minimization (RIPPL) heuristic, which iteratively updates variable-wise penalty weights to better approximate the L0-norm through adaptive, partial penalties. By selectively penalizing small variable magnitudes, RIPPL efficiently promotes sparsity and ensures reliable convergence. This targeted reweighting outperforms standard L1 minimization, enhancing both solution quality and scalability.

Extensive experiments on real-world case studies and simulated networks validate our approach, demonstrating significantly faster runtimes and up to a 74.43% improvement in optimality gaps over state-of-the-art methods. Our solution's scalability enables results within just 3.12% of the optimal solution for large instances.

3 - Data-Driven Behavioral Science for Smarter Transportation Systems

Tianming Liu, University of Michigan, Ann Arbor, MI, United States

Modern transportation systems are rapidly evolving, creating complex challenges that demand a deeper understanding of human behavior. In this talk, I will present my research on developing data-driven and behaviorally-informed solutions for these challenges. My approach integrates diverse data sources with advanced modeling techniques, from econometrics to generative AI, to create fine-grained representations of user decision-making. First, I will discuss my work analyzing large-scale operational data from ride-sourcing platforms. This research uncovered how individuals respond to system incentives and led to the design of novel strategies that significantly enhance system performance and user welfare. Second, I will explain how I've used behavioral queuing theory to understand and mitigate seemingly irrational behaviors, such as excessive driver waiting at terminals, resulting in effective non-pricing management solutions. Finally, I will present my work in pioneering the use of large language models to simulate complex and diverse human travel behavior. This line of research, which bridges AI with behavioral science, is foundational to developing more realistic and adaptable behavior simulation tools for creating efficient, equitable transportation systems.

4 - Friends OR Foes: Do Interference and Staggered Introduction Matter for the Impact of Bike-Sharing on Public Transit?

Yanchao Li, New York University, New York, NY, United States, Daniel Vignon

The rapid expansion of bike-sharing systems has reshaped urban transportation, raising key questions about their interaction with public transit. Accurately evaluating this relationship is critical for guiding infrastructure investment and policy. However, most causal analyses of bike-sharing effects overlook two key challenges: interference—where treatment in one area (e.g., bike station installation) affects outcomes in neighboring areas—and staggered, time-varying treatment, as bike-sharing programs often expand gradually over time. These complexities violate core assumptions in standard causal models, such as SUTVA, potentially leading to biased estimates and misguided policy conclusions.

This study investigates the extent to which interference and staggered treatment assignment influence estimates of bike-sharing's effect on public transit, focusing on New York City's CitiBike program. We: (1) quantify the bias introduced by ignoring spatial spillovers; (2) compare traditional causal estimates with models that account for both interference and treatment timing; and (3) assess how these differences impact infrastructure planning and subsidy allocation.

By applying advanced causal inference methods that explicitly incorporate spatial and temporal complexities, this research provides a more accurate and policy-relevant understanding of bike-sharing's role in urban mobility systems. Findings aim to support more effective coordination between bike-sharing and public transit, maximizing accessibility, efficiency, and investment impact.

SE08

Building A Level 3 A310

Communication and AI

Invited Session

Information Systems

Chair: SIDDHARTH BHATTACHARYA, George Mason University, FAIRFAX, United States

1 - Enhancing Nonprofit Operations with AI Chatbots: the Role of Humanization and Emotion

SIDDHARTH BHATTACHARYA, George Mason University, Fairfax, VA, United States

As nonprofit organizations increasingly adopt artificial intelligence (AI) technologies to sustain engagement and optimize resource allocation, the design of AI-enabled chatbots remains a critical consideration. This study examines the effects of emotional appeals and anthropomorphism in chatbot interactions within a nonprofit context. Through a randomized field experiment at an organization that advocates for minority women's health, we analyze how varying levels of emotional intensity and human-like features influence service completion rates. Our findings reveal an inverted U-shaped relationship between emotional appeals and engagement, wherein moderate emotional intensity enhances service completion, but intense emotional content leads to disengagement. Notably, increased anthropomorphism negatively impacts service completion, suggesting that human-like chatbots may trigger expectancy violations in nonprofit contexts. Furthermore, anthropomorphism moderates the effects of emotional appeals, amplifying their negative impact when emotional intensity is high. These results highlight the complexities of chatbot design in nonprofit contexts, challenging conventional wisdom on persuasive messaging and human-like AI interactions. Our study provides actionable insights for organizations deploying AI-driven engagement tools, emphasizing the need for a balanced approach to emotional framing and anthropomorphic design to optimize user experience and service effectiveness.

2 - When Benevolence Backfires: the Paradoxical Effect of Transparency on Chatbot Donations

Mahmood Farahrooz, University of Wisconsin-Milwaukee, Milwaukee, WI, United States, Scott Schanke, Shweta Singh

As the use of artificial intelligence, particularly chatbots, has increased, so too have privacy concerns. While some believe that transparency could create trust in users and improve engagement, other research shows that it may harm downstream consumption outcomes. This study, in the context of charitable giving, investigates how data disclosure affects donation behavior when people interact with a chatbot. We used a between-subjects experiment (N = 218) and randomly assigned participants to transparent and non-transparent data disclosure conditions and subsequently engaged them on behalf of a charity. Our results show that transparency alone does not have a direct effect on individual donation; instead, transparency greatly impacts an individual's perception of benevolence, which surprisingly, negatively influenced donation, and chatbot likeability. People who were more concerned about privacy were less likely to donate when they were told how their data would be used. These findings highlight the complex and sometimes counterproductive role of transparency in AI-driven charitable interactions.

3 - Minimizing Coordination Costs, Maximizing Care: Clinical Teams Under Emrs Through the Lens of Transaction Cost Economics

Bryan Sullivan, George Mason University, Fairfax, VA, United States

Electronic Medical Records (EMRs) promise to streamline clinical coordination and enhance care quality, yet empirical evidence is mixed, with some studies showing higher costs and clinician burden when vendor differences are ignored. This paper asks: How do Epic and Cerner systems differentially moderate perioperative coordination costs and quality reporting? Drawing on Transaction Cost Economics and Task-Technology Fit, we analyze 15.3 million U.S. anesthesia encounters (2018–2024) linked to CMS quality metrics and facility go-live dates using event-study models and provider-clustered fixed-effects OLS. We find that Cerner implementations reduce missed quality reports by 10 percentage points (versus no change under Epic), attenuate the increase in case duration (3.99 vs. 4.43 minutes), and lower mean professional charges by \$102 per encounter compared to Epic. By unpacking vendor-level EMR effects, we challenge the one-size-fits-all assumption and offer actionable guidance for hospital IT strategy and system selection.

4 - Now You See Me: How Streaming Platforms Enable Diversity in Movies

Naveen Thotadamane Basavaraj, Carnegie Mellon University, Irving, TX, United States

Our research analyzes the degree to which the unique economic characteristics of streaming platforms have allowed for the distribution of movies — in our case movies with diverse casts — that would not have been economically feasible in traditional time- and space-bound physical channels. To do this, we combine several datasets documenting cast diversity in movies, the consumption of movies across theatrical and streaming platforms, and detailed information on the promotion of movies within streaming platforms. Our results show that racial diversity is 16.8% higher for movies made for streaming platforms compared to movies made for theatrical channels. Beyond casting, we apply large language models to extract plot and character archetypes and find that minority actors in streaming movies are more evenly distributed across narrative and character types, suggesting that they are less likely to be typecast into stereotypical roles. We also show that the differences in racial representation between streaming and theatrical channels stem from the structural and algorithmic advantages of streaming platforms, namely unlimited shelf space, data-driven personalization, and recommendation algorithms that promote diverse content to broad audiences. We also find that major social movements like Black Lives Matter are associated with increased consumption but not immediate production of diverse content. Our results highlight the potential for streaming platforms to act as a corrective force against long-standing inequalities in media representation.

SE09

Building A Level 3 A311

Topics in Generative AI

Job Market Showcase

Job Market Showcase

Chair: Anjana Arunkumar, Northeastern University, Oakland, CA, United States

1 - How Do Humans Respond to Creators' Adoption of Generative AI? Insights from the Digital Artwork Market

Yingxin Zhou, Georgia State University, Atlanta, GA, United States, Ling Xue, Arun Rai

As generative artificial intelligence (AI) tools become increasingly prevalent in the creation of artistic content, questions emerge regarding how audiences respond to creators who adopt such technologies. Drawing on data from an online art community, this study examines the impact of artists' generative AI adoption on audience attitudes and explores how these effects may vary based on artists' professional experience. The findings reveal a nuanced relationship between AI adoption and audience perceptions, with distinct patterns across experience levels. This research contributes to ongoing discussions about human responses to AI adoption and offers implications for platforms, creators, and policymakers navigating the evolving boundaries of human-AI collaboration.

2 - Enhancing Security and Fairness in AI-Driven Natural Language Systems

Dorde Klisura, The University of Texas at San Antonio, San Antonio, TX, United States

AI-driven natural language systems increasingly shape decisions in critical sectors such as healthcare, finance, and public administration. However, as these systems embed deeper into sensitive domains, ensuring their security, fairness, and accountability has become essential.

My research addresses these challenges through three interrelated approaches:

- Mitigating dialect bias:** To ensure fair treatment of users from diverse linguistic backgrounds, I developed a multi-agent framework that translates dialect-specific queries into standardized language, significantly reducing performance disparities across over 50 English dialects. This approach achieved an 82% reduction in dialect bias, ensuring equitable access to critical information.
- Securing AI-driven database interfaces:** I discovered zero-knowledge schema inference attacks against Text-to-SQL systems, demonstrating that adversaries can reconstruct hidden database structures merely by querying the system. In response, I designed adaptive defenses that safeguard database schemas while preserving query accuracy, effectively mitigating this critical security vulnerability.
- Permission-aware LLM access control:** I introduced a novel approach embedding explicit access control directly into Large Language Models (LLMs). By integrating user permissions at the prompting and verification stages, the models proactively enforce fine-grained access constraints, achieving 97% accuracy in permission adherence and blocking 94% of unauthorized information disclosures.

Collectively, my research advances AI systems that not only perform efficiently but also uphold stringent standards for security, fairness, and responsible use, ensuring technology empowers and protects all users equally.

3 - Uncovering the Cognitive Mechanisms of Large Language Models in Behavior Modeling

Yu Qin, Arizona State University, Tempe, AZ, United States, Olivia Sheng

Large Language Models (LLMs) are catalyzing a significant paradigm shift in Artificial Intelligence (AI) and reshaping business research through their remarkable proficiency and emulation of complex human-like behaviors. Despite their widespread application, a critical gap exists in understanding the underlying cognitive mechanisms driving these behaviors. Current evaluation methods, primarily observing LLM responses, are insufficient as responses may not faithfully reflect internal knowledge or computational processes. This research proposes to employ Sparse Autoencoders (SAEs) to identify and interpret these latent cognitive features, moving beyond surface-level observations. The principal contribution of this study is to pioneer a deeper, more mechanistic understanding of LLM cognition. By aiming to genuinely uncover the internal features that dictate LLM behaviors, this research provides crucial insights essential for assessing and enhancing their reliability, trustworthiness, safety, and stability – foundational elements for their responsible adoption in high-stakes business environments. This investigation unlocks novel perspectives, opening new frontiers for business research and practice. It equips researchers and practitioners with a more nuanced comprehension, enabling the development of more robust, fair, and transparent AI-driven solutions. Ultimately, by illuminating the cognitive underpinnings of LLMs, this work will foster more informed, effective, and ethical integration of these powerful technologies, significantly advancing how businesses leverage AI for innovation and decision-making.

4 - Modeling Interpretation in Data Visualization: Bridging Computation, Language, and Decision Making

Anjana Arunkumar, Northeastern University, Oakland, CA, United States

As artificial intelligence systems generate, annotate, and mediate ever more of the world's information, a critical question emerges: how do people make sense of it? This research program models the process of interpretation itself—how humans extract meaning from visualizations and language, especially when navigating complex or unfamiliar data.

Situated at the intersection of data visualization, cognitive science, and natural language processing, the work combines empirical studies with interactive system design to examine how users perceive, recall, and reason with data. It introduces modeling frameworks that capture how annotation strategies, linguistic framing, and individual differences shape understanding, trust, and decision quality. These insights inform the creation of human-in-the-loop tools that support generalization, reduce bias, and make AI-driven outputs more transparent and context-aware.

Rather than isolating visualization or language as endpoints, this research treats them as generative interfaces—sites where human and machine interpretations converge. Applications span multilingual visualization, instruction tuning, and decision support, with a shared emphasis on adaptive systems that respond to diverse user needs and cognitive constraints. The result is a design and modeling agenda for building interpretable, equitable systems—tools that go beyond generating outputs to actively support how people think, reason, and decide with data.

SE10

Building A Level 3 A312

Generative AI in Finance

Invited Session

Finance

Chair: Renyuan Xu, Stanford University, Palo Alto, CA, United States

Co-Chair: Haoyang Cao, Johns Hopkins University, Baltimore, MD, United States

1 - Financial TIME Series Forecasting via Generative Adversarial Networks

Mihai Cucuringu, UCLA, Los Angeles, CA, United States

We investigate the use of Generative Adversarial Networks (GANs) for probabilistic forecasting of financial time series. To this end, we introduce a novel economics-driven loss function for the generator. This newly designed loss function renders GANs more suitable for a classification task, and places them into a supervised learning setting, whilst producing full conditional probability distributions of price returns given previous historical values. Our approach moves beyond the point estimates traditionally employed in the forecasting literature, and allows for uncertainty estimates. Numerical experiments on equity data showcase the effectiveness of our proposed methodology, which achieves higher Sharpe Ratios compared to classical supervised learning models, such as LSTMs and ARIMA. Finally, we discuss extensions to the multi-asset setting by leveraging cross-asset relationships through a graph-based probabilistic ensemble framework.

2 - Training Foundation Models for Algorithmic Trading

Atlas Wang, XTX Markets & UT Austin, New York, NY, United States

In this talk, I will introduce the pioneering efforts of XTX Markets in applying large-scale foundation models to algorithmic trading. I will highlight both the unique challenges and exciting opportunities in forecasting financial markets using sophisticated deep neural networks trained on massive, heterogeneous datasets. Specifically, I will present a high-level discussion how our models can be trained to utilize diverse market data to generate robust forecasts across thousands of financial instruments globally. I will also cover our experiences in "neural scaling laws" on market data as a novel modality, amid extreme market noise and volatility. Finally, I will introduce our latest research initiatives, emphasizing XTX's substantial investments in computational infrastructure, innovative modeling approaches, and proprietary research.

3 - Diffusion Models for Sequential (Financial) Data Generation

Yinbin Han, New York University, New York, NY, United States, Haoyang Cao, Minshuo Chen, Renyuan Xu

We propose a novel paradigm for sequential data generation based on diffusion models. To ensure the adaptedness of the generated sequences, we introduce an enlarged filtration that aligns with the spatial-temporal structure of the data. By decomposing the generation process in the spirit of next-token prediction, our framework effectively incorporates diffusion models for sequential synthesis while preserving training efficiency. We further establish theoretical guarantees, including universal approximation, score estimation, and distribution estimation, while addressing the technical challenges posed by unbounded domains. Numerical experiments demonstrate the effectiveness and efficiency of the proposed approach.

4 - Diffusion Factor Models: Generating High-Dimensional Returns with Factor Structure

Yumin Xu, Peking University, Beijing, China, People's Republic of, Minshuo Chen, Renyuan Xu, Ruixun Zhang

Financial scenario simulation is essential for risk management and portfolio optimization, yet it remains challenging especially in high-dimensional and small data settings common in finance. We propose a *diffusion factor model* that integrates latent factor structure into generative diffusion processes, bridging econometrics with modern generative AI to address the challenges of the curse of dimensionality and data scarcity in financial simulation. By exploiting the low-dimensional factor structure inherent in asset returns, we decompose the score function—a key component in diffusion models—using time-varying orthogonal projections, and this decomposition is incorporated into the design of neural network architectures. We derive rigorous statistical guarantees, establishing non-asymptotic error bounds for both score estimation at $O(d^{5/2} n^{-2/(k+5)})$ and generated distribution at $O(d^{5/4} n^{-1/(2k+10)})$, primarily driven by the intrinsic factor dimension k rather than the number of assets d , surpassing the dimension-dependent limits in the classical nonparametric statistics literature and making the framework viable for markets with thousands of assets. Numerical studies confirm superior performance in latent subspace recovery under small data regimes. Empirical analysis demonstrates the economic significance of our framework in constructing mean-variance optimal portfolios and factor portfolios. This work presents the first theoretical integration of factor structure with diffusion models, offering a principled approach for high-dimensional financial simulation with limited data.

5 - Robust Financial Modeling: a Principled Framework for Generative Data Augmentation

Amine Aboussalah, New York University, New York, NY, United States

Effective machine learning for financial modeling is often challenged by limited data, leading to overfitting and poor out-of-distribution generalization. This talk introduces a principled framework for data augmentation to address these critical challenges. This framework models diverse data structures, from time series to graph-

structured networks. We establish a theoretical foundation for our framework using Rademacher complexity to offer a regret bound on the generalization error, thereby guaranteeing performance under specific conditions. Through empirical evaluation on diverse datasets, we demonstrate that our approach not only significantly outperforms state-of-the-art augmentation techniques but also offers superior computational efficiency. This research provides a robust, theoretically-grounded solution for enhancing the performance and reliability of generative AI models in real-world applications such as finance.

SE11

Building A Level 3 A313

Data-driven Healthcare Management

Job Market Showcase

Job Market Showcase

Chair: Haya Alshayji, Pennsylvania State University, State College, PA, United States

1 - Interpretable Design of Multi-Agent Reinforcement Learning In Mobile Health Applications

Hossein Hejazian, Harvard University, Cambridge, MA, United States, Mohammad Zhalechian, Soroush Saghaian

Multi-agent mHealth strategies can effectively address multifaceted health goals by coordinating multiple interventions. Existing micro-randomized trials are single-agent and lack data on agent interactions. Reinforcement learning offers an online framework to optimize multi-agent policies but faces engagement, exploration-exploitation, and interpretability challenges. We provide novel solutions for holistic mobile health interventions by proposing an algorithm designed to deliver personalized interventions to users with the aim of forming an overall health habit. We introduce the multi-agent analytical framework, where decisions are made by an automated manager as well as several agents. In this framework, each agent is responsible for one aspect of the user's health (e.g., sedentary behavior, diet, or physical activity). The manager's role is to coordinate the agents such that the system achieves its goal. Our framework sequentially determines the agent the manager should select at each interaction with a user, as well as the intervention taken by the designated agent. We develop a holistic framework with a provable performance guarantee which provides a data-driven policy in our setting. Our results provide insights into how such a holistic framework helps form healthy habits for users by taking advantage of mobile health technologies.

2 - Identifying and Predicting Vaccine Hesitancy Using Data-Driven Behavioral Modeling

Citina Liang, University of Southern California, los angeles, CA, United States, Sze-chuan Suen

Understanding vaccine uptake behavior is critical to effective vaccination deployment, especially for diseases like COVID-19 where both the rate and timing of vaccination can influence transmission dynamics. Despite widespread vaccination campaigns, challenges remain in promoting vaccine uptake effectively due to limited insight into population subgroups with varying attitudes and behaviors.

In this work, we use survey and surveillance data to identify population subgroups with distinct vaccine uptake behaviors and predict vaccination timing using demographic characteristics and self-reported attitudes. We leverage longitudinal survey data capturing individual beliefs, behaviors, and demographics, along with vaccination registry records. These datasets present distinct information on the population and present unique challenges. The longitudinal survey data contain non-uniformly timed attitudinal measures, with key questions asked at different waves, leading to missing or inconsistent responses across individuals. Meanwhile, the vaccination registry data offer basic demographics but lack behavioral context, limiting direct insight into vaccine decision-making.

To address this, we apply latent class analysis and unsupervised machine learning methods such as k-means to identify evolving subgroups of vaccine hesitancy. We then use supervised learning models, including regression and gradient boosting classifiers, to predict both likelihood and timing of vaccine uptake based on subgroup membership and individual-level features. Our findings reveal meaningful behavioral subgroups and predictive patterns that support targeted public health interventions. This data-driven approach enhances the precision of vaccination strategies, particularly in diverse populations such as those in Los Angeles County.

Acknowledgments: This research was supported by the Merck Investigator Studies Program (MISP), Merck & Co., Inc.

3 - Early Prediction of Patient Admission Likelihood in Emergency Departments

Farzad Zeinali, Clemson University, Clemson, SC, United States

Emergency Department (ED) overcrowding remains a persistent challenge in U.S. hospitals, driven largely by admitted patients who wait in the ED for inpatient beds. This study aims to predict the likelihood of patient admission during the early stages of an ED visit, particularly before physician evaluation. Anticipating inpatient demand early in the care process is a critical step toward operational planning and improving patient flow between the ED and other hospital departments. We developed a machine learning (ML) framework that integrates both structured data and unstructured triage notes to support real-time decision-making. The dataset includes demographic, clinical, and visit-related features. Importantly, we identified impactful variables not available in tabular format and extracted them from free-text triage notes using open-source large language models (LLMs). In total, over 30 variables were used to train and evaluate multiple state-of-the-art classification algorithms. By identifying the top-performing models and the most influential predictors, our analysis shows that accurate prediction of admissions is feasible early in the ED process. These findings demonstrate the potential of data-driven decision support systems to enhance coordination between the ED and inpatient units, improve operational responsiveness, and enable proactive interventions that reduce boarding delays and alleviate ED overcrowding.

4 - Integrating AI and Multilevel Modeling to Study Sdoh Impacts on Diseases of Despair

Haya Alshayji, Pennsylvania State University, State College, PA, United States

This talk explores integrating AI-driven methods and multilevel statistical modeling to examine the impact of Social Determinants of Health (SDoH) on Diseases of Despair (DoD). This research uses hierarchical models and machine learning to enhance risk prediction and inform targeted interventions to address health disparities in vulnerable populations.

SE12

Building A Level 3 A314

Models of Contracts and Compensation

Job Market Showcase

Job Market Showcase

Chair: Kathleen Tomon, Colorado School of Mines, Lakewood, CO, United States

1 - Combating Overbilling in Outsourced Projects: a Dynamic Auditing Mechanism

Like Bu, The University of Texas at Dallas, Richardson, TX, United States, Shivam Gupta, Milind Dawande, GANESH JANAKIRAMAN

Firms routinely outsource their business requirements to external agents for many reasons; e.g., to focus on their core competencies or to save costs. However, overbilling by agents has been well-acknowledged as a notorious problem across major industries, including healthcare, information technology, legal services, engineering, and consulting. Mitigating overbilling is challenging in practice; typical options for firms include auditing agents or suspending relationships with them. While firms would like to combat overbilling by utilizing these options, they also want to minimize their total cost, including agents' payments and auditing cost. We consider a repeated

principal-agent setting in which, in each period, the principal makes two decisions: (i) Whether to allocate the task to the agent or an outside option. (ii) If the agent executes the task, whether to audit the agent's cost. We propose a class of *Dynamic Auditing* mechanisms under which it is near-optimal for the agent to report his costs truthfully and the principal's cost is also near-optimal, for sufficiently large discount factors. We are the first to investigate the role of auditing in repeated settings – we show that auditing is an effective tool for the principal if either (i) the unit cost of auditing is sufficiently small, or (ii) the principal lacks knowledge of the agent's cost distribution. We also show that our results are robust when there is competition among multiple agents for the principal's business.

2 - Segmentation and Market-Based Pricing for Uncertain Episodic Service Demand: Case Study of Truckload Service Procurement for Fema

Shraddha Rana, INSEAD, Fontainebleau, France

Following natural disasters and man-made disruptions, the uncertain, episodic, and time-sensitive demand for relief services makes it difficult for humanitarian actors to compete with private-sector customers in securing commercial service capacity, leading to high costs without guarantee of higher service levels. Our aim is to strategically design and implement flexible contracts that can help humanitarian organizations get the relief services on time and economically. We study a case of transportation service procurement for the US Federal Emergency Management Agency (FEMA). Most disaster management literature focus on the procurement, allocation, and routing of the relief supplies. However, it is imperative to first ensure that transportation capacity would be available to move the supplies. We contribute to the disaster management, truckload procurement, and dynamic-pricing literature by i) studying transportation service procurement for FEMA, without capacity guarantee; ii) considering uncertain episodic demand in response to disasters; iii) modeling a market-based price contract; and iv) examining the case of dynamic-pricing where the buyer, i.e., FEMA offers the price as opposed to the seller, i.e., transportation service providers. We propose a segmentation framework that leverages the certainty of location of supply nodes and general timing of weather-based disasters, while also allowing flexibility of real-time pricing. We find that on implementing the proposed segmentation and market-based pricing, FEMA would have saved \$9.85 million in transportation service procurement costs in 2 years while still paying 71% over market rates leading to high service levels.

3 - Income Pools for Superstar Markets

Craig Fernandes, University of Toronto, Toronto, ON, Canada, Timothy Chan, Ningyuan Chen

"Superstar" markets, where a small portion of individuals earn disproportionately high incomes, are common in fields like entrepreneurship, sports, and entertainment. Participants in these markets face significant income uncertainty, which can deter entry or prompt early exit. To address this difficulty, we propose income pools -- contracts where individuals agree to share a portion of future earnings with pool members if a specific salary milestone is achieved. To date, hundreds of income pool contracts have been signed in practice. This paper develops the first mathematical model to analyze such contracts, focusing on stability (i.e., pools where no agents leave or join). When constrained to creating a single pool, we show that a stable pool always exists with specific structural properties. Generally, stronger agents tend to be more "collaborative" and favor larger pools, while weaker agents are more "selfish" and prefer smaller pools. Next, we analyze pools that adhere to a maximum size (mirroring current practice) and show that stability persists. These pools typically require more weaker agents than strong ones and we find an interesting "Pareto dominance" result, whereby all agents in a stable pool prefer a particular unique stable pool. Finally, we study general partition structures, prove that a stable partition always exists under certain conditions, and provide an algorithm to construct such a partition. We conclude with a case study on data from 2000 professional baseball players to demonstrate a 20%-30% increase in social welfare if players join income pools, under varying contract parameters.

4 - Profit Maximization and Compensation Stability in Service-Oriented Businesses

Kathleen Tomon, Colorado School of Mines, Golden, CO, United States, Alexandra Newman

Service-oriented businesses rely heavily on labor and expertise, making compensation structures critical to their operational success. Various commission structures exist, but current research lacks effective methods to optimize these systems in service-oriented settings. We present a generalized optimization model that revises existing commission and bonus scales to: (i) ensure target business gross profits, (ii) minimize deviation from prior employee compensation levels, (iii) institute a service price increase, and (iv) maintain internal consistency across bonus tiers. Applied to real-world data from a service-oriented business in a major U.S. metropolitan area, our model improves average gross profit per commission tier by 29%, while preserving compensation equity and outperforming current methods.

SE13

Building A Level 3 A315

Optimization Under Uncertainty I

Job Market Showcase

Job Market Showcase

Chair: Yizhe Huang, University of Texas at Austin, Austin, TX, United States

1 - Robust Decision-Making in Partially Observable Markov Decision Processes

Tong Li, University of Houston, Houston, TX, United States

We study robust decision-making in partially observable Markov decision processes (POMDPs) under two key challenges: distributional uncertainty and adversarial perturbations. First, we develop a distributionally robust POMDP framework using distance-based ambiguity sets to model uncertainty in rewards and transitions. We prove value function convexity and derive tractable reformulations for commonly used statistical distances. Second, we address state adversarial perturbations by proposing a risk-aware certification approach based on exponential utility. This method quantifies the lower bound of cumulative reward under adversarial noise and risk-sensitive evaluation. Experiments on OpenAI Gym and a machine replacement problem show that distance-based robustness improves out-of-sample performance, while risk-averse training enhances certified policy performance under state perturbations. These complementary approaches together advance the robustness of POMDPs in adversarial and uncertain environments.

2 - Stochastic Dual Dynamic Programming (SDDP) for Decision Making Under Conflict and Uncertainty

Benjamin Hamlin, Clemson University, Clemson, SC, United States, Cheng Guo

We introduce a Stochastic Dual Dynamic Programming method for multiobjective or single-objective-parametric multistage stochastic programming problems. A parametric approach eliminates the bottleneck caused by the Simplex method allowing for only two objectives. Commercial solvers are used to efficiently perform forward and backward passes. We demonstrate the algorithm on a triobjective Brazilian hydrothermal problem.

3 - A Deep Learning Approach to Multistage Stochastic Programming

Yizhe Huang, University of Texas at Austin, Austin, TX, United States

Multistage stochastic programming (MSP) is a powerful framework for sequential decision-making under uncertainty with complex constraints, but solving large-scale instances remains challenging due to the curse of horizon and nested optimization structure. Existing value-based methods approximate the cost-to-go function recursively, but suffer from compounding approximation errors, require solving constrained subproblems at every stage, and lack direct policy representations. In this work, we propose a principled policy-based alternative that parameterizes decision rules with neural networks and optimizes them via an augmented Lagrangian minimax formulation. Our single-loop stochastic gradient algorithm avoids nested subproblem solvers and handles constraints through dual updates, enabling scalable learning in constrained MSPs. We provide theoretical guarantees on feasibility, optimality, and generalization, and demonstrate superior empirical performance over value-function methods such as DVFN and SDDP across high-dimensional benchmark problems in energy and production planning.

SE14

Building A Level 3 A316

Algorithmic Fairness and Equity in Operations

Contributed Session

Contributed

Chair: Nowshin Sharmile, North Carolina A&T State University, 3533 Ivy Commons Drive, Raleigh, NC, 27606, United States

1 - Bias in AI Hiring: Assessing Fairness in Algorithmic Resume Screening

Xiahua Wei, University of Washington, Bothell School of Business, Bothell, WA, United States, Naveen Kumar

The rapid adoption of AI in hiring processes, particularly in resume screening, raises pressing concerns about fairness and bias. This study investigates the potential biases in large language models (LLMs) by analyzing Generative AI's evaluation of resumes that vary systematically across legally protected attributes (such as race, color, gender, age, and disability status) and non-protected attributes (such as employment gap, education, and writing quality), both of which can trigger explicit and implicit biases. Drawing from prior research on employment discrimination, we assess how Generative AI responds to these variations and the potential heterogeneity in biases. While our findings show limited evidence of bias for some protected attributes, results for others are mixed, especially when protected and non-protected attributes appear together. These results demonstrate the complexity of detecting and interpreting bias in AI systems. Our work contributes to ongoing efforts to promote fairness in algorithmic hiring and highlights the need for scrutiny of AI tools to support equitable employment practices.

2 - Fair to Fare Well: How Algorithmic Fairness Benefits Firms on a Platform?

Yaping Li, College of Management and Economics, Tianjin University, Tianjin, China, People's Republic of, Yunchuan Liu, Zhe Yang, Jianxiong Zhang

In this paper, we examine the impact of algorithmic fairness—through unawareness (e.g., equal treatment) or awareness (e.g., equal opportunity)—on firms that rely on platform algorithms for decision-making. Specifically, in industries such as lending and hiring, lenders and employers use platform algorithms to predict candidates' qualifications. We develop a game-theoretic model involving a platform and a firm, the platform can invest in learning to improve algorithm accuracy for better prediction of candidates' qualifications and the firm can decide which candidates to accept by setting score thresholds, whereas the candidates belong to either a regular group or a protected group that is biased against by platform algorithms. Counterintuitively, we show that algorithmic fairness, while seemingly detrimental to firms' profitability due to the constraints it imposes, can actually benefit firms. Interestingly, we find that algorithmic fairness can hurt the protected group or benefit the regular group. Our study also shows that whether equal opportunity or equal treatment outperforms in maximizing firms' profits depends on the benefit-loss ratio. When this ratio is sufficiently high, the firm has higher profits under equal opportunity; otherwise, the firm has higher profits under equal treatment. This study also provides implications for policymakers on when to allow or prohibit the use of sensitive attributes in platform algorithms.

3 - On Fairness of Process Flexibility

Lyu Hongyao, Shanghai Jiaotong University, Shanghai, China, People's Republic of, Hanzhang Qin, Huan Zheng

In the process flexibility literature, it is well known that "a little flexibility goes a long way" (Jordan and Graves 1995). This flexibility folklore states that a sparse process flexibility structure can capture nearly all the benefits of full flexibility regarding demand satisfaction and/or capacity utilization. While the efficiency metrics have been well assessed in the literature, the performance of sparse structures concerning fairness measures remains under-explored. Fairness is a crucial objective in resource allocation, driven by companies' social responsibility and the needs of humanitarian operations. In this study, we investigate the worst-case performance of sparse flexibility structures regarding the max-min demand fill rate. Through primal-dual analysis, we show that the max-min fill rate of a structure is determined by the demand nodes subset with the minimum ratio of neighboring capacity to demand. We further characterize the worst-case max-min fill rate of k -chain and Ψ -expander structures. The results suggest that a more connected structure is needed to achieve outcomes close to full flexibility. Additionally, we demonstrate that there exists an allocation that can achieve both efficiency (i.e., max-flow) and fairness (i.e., max-min fill rate). Based on our theoretical findings, we propose a computationally efficient heuristic, the Cover Ratio heuristic, to generate effective flexibility structures. Our theoretical analysis also extends to the min-max shortfall metric. We validate our theoretical findings using real-world data from the Texas Department of State Health Services and conclude that incorporating flexibility structure design can substantially enhance both fairness and efficiency in vaccine allocation.

4 - Designing Truthful Mechanisms for Asymptotic Fair Division

Yuang Shen, UIUC, Urbana, IL, United States, Jugal Garg, Vishnu V Narayan

We study the problem of fairly allocating a set of m goods among n agents in the asymptotic setting, where each item's value for each agent is drawn from an underlying joint distribution. Prior works have shown that if this distribution is well-behaved, then an envy-free allocation exists with high probability when $m = \Omega(n \log(n))$. Under the stronger assumption that item values are independently and identically distributed (i.i.d.) across agents, this requirement improves to $m = \Omega(n \log(n)/\log(\log(n)))$, which is tight. However, these results rely on non-strategyproof mechanisms, such as maximum-welfare allocation or the round-robin algorithm, limiting their applicability in settings with strategic agents.

In this work, we extend the theory to a broad and more realistic class of joint value distributions, allowing for correlations between agents, atomicity, and unequal probabilities of having the highest value for an item. We show that envy-free allocations continue to exist with high probability when $m = \Omega(n \log(n))$. More importantly, we give a new randomized mechanism that is truthful in expectation, efficiently implementable in polynomial time, and outputs envy-free allocations with high probability. We further extend our mechanism to the settings of asymptotic weighted fair division, and multiple agent types and good types, proving new results in each of these cases.

5 - Improving Fairness of Spatiotemporal Travel Demand Modeling with Dynamic Weighting

Xiaojuan Zhang, University of Florida, Gainesville, FL, United States, Xilei Zhao

Fairness has been a critical concern in travel demand modeling, as inequitable predictions can lead to disproportionate service allocation and reinforce existing mobility disadvantages. Recent studies have proposed methods to improve the fairness of demand forecasts, particularly focusing on reducing spatial disparities such as unequal prediction accuracy across different neighborhoods. However, fairness over time has received far less attention, despite being equally important. Inaccurate and unfair predictions during certain times of day, especially peak commuting hours, can result in misallocated transportation resources and severe service mismatches, causing delays and reduced service quality for those who depend on timely and reliable mobility options. This study addresses this challenge by proposing a novel dynamic weighting methodology called Fair Focal Loss. Specifically, Fair Focal Loss can dynamically learn and adjust the weights of different losses based on each loss's performance throughout the course of training. This mechanism can guide the travel demand models to prioritize time periods and locations where fairness improvements are most needed. We apply the methodology to real-world ridesourcing-trip data in Chicago, framing predictions at different hours of the day as separate losses. Results show that the proposed method significantly improves the fairness of travel demand prediction at both spatial and temporal scales. Also, we find that the dynamic-weighting mechanism balances the learning process by driving the model to prioritize peak-hour predictions over off-peak hours. Overall, this study offers transportation professionals a powerful tool to enhance fairness in spatiotemporal travel demand modeling for better real-time traffic management and resource allocations.

6 - Optimizing Weight-Based and Nutritional Fairness Metrics in Charitable Food Distribution

Nowshin Sharmile, North Carolina A&T State University, Greensboro, NC, United States, Lauren Davis

Food insecurity—defined as limited or uncertain access to sufficient food for a healthy, active life—continues to be a major concern in the United States. In 2023, 13.5% of households, representing 47.4 million people, faced food insecurity. Food banks and nonprofit organizations play a vital role in supporting these communities by distributing donated food. Traditionally, their performance has been measured by the total pounds of food delivered. However, this approach often overlooks the nutritional quality of the food provided. This study focuses on the Food Bank of Central and Eastern North Carolina (FBCENC) and applies the Healthy Eating Research (HER) Nutrition Guidelines to assess distribution fairness from both a weight and nutritional perspective. We classified distributed food into tiers based on HER criteria and built a multi-branch, multi-agency network optimization model. The model incorporates two key objectives: maximizing fairness in agencies' serving capacities and minimizing shortfalls in the distribution of nutritious food. A lexicographic approach is used to explore how prioritizing one objective affects the other. Our findings highlight the trade-off: focusing solely on weight improves overall coverage but can worsen nutritional disparities, while prioritizing nutrition may limit the total amount

of food served. The lexicographic method offers a middle ground, enhancing nutritional equity while maintaining reasonable service levels. These results point to the importance of integrated, multi-objective strategies for food banks aiming to ensure both fair and health-conscious food distribution.

SE15

Building A Level 4 A401

Operational Resilience in Aviation

Invited Session

Air Transportation Section

Chair: Alexander Estes, University of Maryland, College Park, MD, United States

1 - Stress-Testing Nas Resilience with Disruptive Scenario Generation

Sinan Abdulhak, University of Michigan, Ann Arbor, MI, United States

Bolstering stakeholders' (e.g., Traffic Managers) ability to respond to common and extraordinary disruptive scenarios is key to national airspace system (NAS) resiliency. Training is a vital tool to building the specialized skills needed to respond to disruptive (e.g., severe weather) events by exposing trainees to commonly encountered scenarios in a low-stakes environment. However, training on historical data neglects the novelty of future events. Furthermore, generating training scripts is highly time consuming, requiring hours of specialized expertise and the need to account for several courses of action, which compounds script writing requirements. We propose generating novel training scenarios that draw upon structural similarities of historical events, and display them intuitively in formats with operational meaning to facilitate immersive training exercises. To accomplish this, we use stochastic optimization models to rigorously generate parameters of synthetic, plausible disruptive scenarios. We then translate these quantitative parameters into a detailed training script using a retrieval-augmented generation (RAG) pipeline built on the GPT-4o large language model (LLM), displayed to stakeholders in a format that carries operational significance. To understand how training scenarios should be displayed, we conduct over 14 stakeholder interviews with Federal Aviation Administration (FAA) Traffic Managers, airport operators, and airline ATM teams. Preliminary results include an LLM-based web app that generates training scripts using key disruptive parameters from the optimization model for the major US 30 airports.

2 - Exploring Multimodal Air Transportation Network Design Through Intelligent Inference

Micah Borrero, University of Michigan, Ann Arbor, MI, United States, Max Li

In the context of short, regional routes, the use of relatively low-capacity vehicles is a practical and often necessary consideration. Preliminary studies suggest that integrating multiple modes of transportation offers distinct advantages, particularly in mitigating congestion associated with the hub-and-spoke structure of conventional air transportation networks. To determine the optimal deployment scenarios for such vehicles, it is essential to have accurate knowledge of origin-destination (OD) demand and true passenger flows. However, publicly available datasets commonly used to estimate air transportation demand are limited, often exhibiting inconsistencies between OD pairs and actual passenger flow data. These inconsistencies are particularly pronounced at the regional level, introducing stochasticity into demand estimates that undermines the interpretability of network construction models and reduces confidence in the resulting deployment strategies. To overcome these limitations, we propose a unified methodology that leverages techniques rooted in Reinforcement Learning to model the stochastic nature of estimated demand within the network design process. Specifically, we simultaneously infer a complete demand dataset while capturing its inherent stochasticity, allowing uncertainty to be directly incorporated into the decision-making process. This enables the development of deployment strategies that are resilient to variability in demand, supporting more reliable and adaptive planning of multimodal air transportation systems.

3 - Pathfinders in the Sky: Formal Decision-Making Models for Collaborative Air Traffic Control in Convective Weather

Jimin Choi, University of Michigan, Ann Arbor, MI, United States, Max Li

Convective weather remains one of the most disruptive factors in air traffic operations, often leading to significant delays and rerouting. A practical response employed by controllers is the use of pathfinder, which is a flight that tests whether weather-affected routes are once again viable. Despite their operational relevance, the underlying decision mechanisms for pathfinding have not been systematically studied. In this talk, we introduce formal decision models that support resilient airspace recovery through collaborative pathfinder operations. Using a Markov chain framework, we capture transitions between operational states such as flight selection and airspace status. Additional models represent flight-level decisions driven by utility trade-offs, as well as offer allocation strategies for pathfinder designation. We also perform an all-rejection scenario analysis to assess systemic vulnerabilities, highlighting how coordination and behavioral factors affect overall resilience. Analysis of data from the US Federal Aviation Administration provides real-world insights into the impact of pathfinder operations and helps guide future model adjustments.

4 - Identification and Case Studies of Operational Disruptions in the U.S. National Airspace System

Jing Xu, University of California, Berkeley, Oakland, CA, United States

Disruptions in the National Airspace System (NAS) lead to significant losses to air traffic system participants and raise public concerns. We apply two methods, cluster analysis and anomaly detection models, to identify operational disruptions with geographical patterns in the NAS since 2010. We identify four types and twelve categories of days of operations, distinguished according to air traffic system operational performance and geographical patterns of disruptions. Anomaly detection results show good agreement with cluster results and further distinguish days in the same cluster by severity of disruptions. Results show an increasing trend in frequency of disruptions, especially post-COVID. Additionally, disruptions happen most frequently in the summer and winter.

5 - Generation of Synthetic National Airspace System Weather Disruptions

Yuhan Miao, University of Maryland, College Park, MD, United States

We develop a data-driven method for developing synthetic weather disruptions. Such disruptions could be used for training of operations managers at airlines or air navigation service providers, or could be used as stress-tests to evaluate national airspace system resilience. Our method combines deep machine learning methods with a heuristic optimization approach.

SE16

Building A Level 4 A402

Optimization Meets Generative AI: Applications and Insights from the Industry/Academia

Invited Session

The Practice Section of Informs

Chair: Konstantina Mellou, Microsoft Research, Redmond, WA, United States

Co-Chair: Burcin Bozkaya, NVIDIA Corporation, Sarasota

1 - Hybrid Metaheuristic Framework for Large-Scale Production Scheduling Using LLM-Generated Heuristics

Amit Kumar, Lyric.tech, Luxembourg, Luxembourg, Sarang Jagdale

Production Planning and Detailed Scheduling (PPDS) presents an NP-hard optimization challenge requiring simultaneous lot-sizing and sequence-dependent changeover optimization across multi-line systems. Traditional mixed-integer programming (MIP) approaches in industrial settings (e.g., >100 SKUs, 50+ periods) suffer from scalability limitations due to solution space fragmentation. We propose a hybrid framework integrating large language model (LLM)-generated heuristics with mathematical optimization to address computational intractability. Our methodology combines: 1) A constraint-guided greedy heuristic (developed via iterative LLM interactions with targeted problem context) achieving 96.2% demand fulfillment in 300 seconds, and 2) A sliding window neighborhood search reducing cumulative changeovers by 29.8% through symmetry-aware permutation analysis. The heuristic solution warmstarts MIP solvers, enabling optimal solutions within 15-20 minutes (vs. >4 hours without initialization) while maintaining a 1.4% average optimality gap. Key innovations include the LLM's synthesis of domain-specific neighborhood

structures that exploit temporal and product-based symmetries invisible to standard branching rules. Computational tests demonstrate resolution of 83% previously unsolvable industrial instances. The framework's effectiveness stems from decoupling lot-sizing feasibility (handled via heuristic search) from changeover optimization (intensified through local search), overcoming MIP's tight constraint coupling. This work establishes a paradigm shift for high-dimensional scheduling by combining generative AI's pattern recognition with operations research's exact methods, particularly where fragmented solution spaces hinder traditional approaches. Practical implications include scalable tools for multi-line capacity balancing and changeover cost minimization in process industries.

2 - Optiguide: Genai for Optimization and Decision Intelligence

Konstantina Mellou, Microsoft Research, Redmond, WA, United States

In this talk, we will describe how LLMs and optimization techniques are leveraged in Microsoft's cloud supply chain to improve efficiency, give insights to planners and answer what-if questions. More broadly, we will discuss ongoing research and future directions in the intersection of operations research and AI.

3 - Solving Supply Chain Planning Problems with Cuopt in Agentic Workflows

Burcin Bozkaya, NVIDIA Corporation, Sarasota, FL, United States

Supply Chain Planning and Management involves various players (eg. suppliers, factories, distribution centers, retail outlets) that require major coordination and complex set of inter-connected decisions. This leads to very large and complex optimization models, which are a major challenge for organizations operating at a global scale who need to build, maintain, solve, interpret and implement them. Add disruption management to this picture, and the complexity multiplies. LLMs and Generative AI has introduced in recent years ability to understand a given set of instructions or modeling constructs, which in turn helps in building, modifying and interpreting complex optimization models. Combined with an agentic AI workflow and tool-calling, Generative AI-assisted end-to-end optimization workflows can now be built and deployed. In this talk, we present NVIDIA's approach and framework to LLM-assisted supply chain planning and optimization in an agentic workflow. We employ cuOpt, NVIDIA's decision optimization library for solving a supply chain planning problem and build an agentic workflow around it. The workflow takes inputs from the user to build and solve an optimization model and generates natural language explanation of the optimization results. The user can then ask questions to explain parts of the resulting solution. The user can further conduct near-realtime what-if analysis by sending prompts to the agent that translates them into model updates and re-solves the problem. We implemented this framework for various business cases and achieved favorable feedback towards explainability of complex models, also providing near-real-time answers via the accelerated cuOpt engine in the context of disruption management.

4 - Learning-Guided Rolling Horizon Optimization for Long-Horizon Flexible Job Shop Scheduling

Cathy Wu, Massachusetts Institute of Technology, Cambridge, MA, United States

Long-horizon combinatorial optimization problems (COPs), such as the Flexible Job-Shop Scheduling Problem (FJSP), often involve complex, interdependent decisions over extended time frames, posing significant challenges for existing solvers. While Rolling Horizon Optimization (RHO) addresses this by decomposing problems into overlapping shorter-horizon subproblems, such overlap often involves redundant computations. In this paper, we present L-RHO, the first learning-guided RHO framework for COPs. L-RHO employs a neural network to intelligently fix variables that in hindsight did not need to be re-optimized, resulting in smaller and thus easier-to-solve subproblems. For FJSP, this means identifying operations with unchanged machine assignments between consecutive subproblems. Applied to FJSP, L-RHO accelerates RHO by up to 54% while significantly improving solution quality, outperforming other heuristic and learning-based baselines. We also provide in-depth discussions and verify the desirable adaptability and generalization of L-RHO across numerous FJSP variates, distributions, online scenarios and benchmark instances. Moreover, we provide a theoretical analysis to elucidate the conditions under which learning is beneficial.

5 - Causal LLM Routing: End-to-End Regret Minimization from Observational Data

Wei Sun, IBM Research, Yorktown Heights, NY, United States, ASTERIOS TSIOURVAS, Georgia Perakis

LLM routing aims to select the most appropriate model for each query, balancing competing performance metrics such as accuracy and cost across a pool of language models. Prior approaches typically adopt a decoupled strategy, where the metrics are first predicted and the model is then selected based on these estimates. This setup is prone to compounding errors and often relies on full-feedback data, where each query is evaluated by all candidate models, which is costly to obtain and maintain in practice. In contrast, we learn from observational data, which records only the outcome of the model actually deployed. We propose a causal end-to-end framework that learns routing policies by minimizing decision-making regret from observational data. To enable efficient optimization, we introduce two theoretically grounded surrogate objectives: a classification-based upper bound, and a softmax-weighted regret approximation shown to recover the optimal policy at convergence. We further extend our framework to handle heterogeneous cost preferences via an interval-conditioned architecture. Experiments on public benchmarks show that our method outperforms existing baselines, achieving state-of-the-art performance across different embedding models.

6 - Llm-Enhanced Flexible and Dynamic Analyses for Model-Based Decision Optimization

Mohsen Bahrami, Massachusetts Institute of Technology, Cambridge, MA, United States

This work introduces an innovative approach to enhance business intelligence (BI) platforms using Large Language Models (LLMs) for flexible and dynamic analyses in model-based decision optimization. Traditional BI platforms often face limitations due to rigid structures. This new system integrates an LLM agent with a user interface, fine-tuned on relevant models, code examples, and data documentation.

The LLM dynamically interprets user requests, selecting appropriate predictive models and optimization techniques. For example, when estimating a new store's market share, the LLM selects and justifies model choices and employs suitable optimization methods such as Ordinary Least Squares (OLS), Generalized Linear Models (GLM), or Particle Swarm Optimization (PSO). This system allows users to define complex analytical tasks through natural language, without needing to write code.

Most important features of this design include dynamic model selection, code generation, real-time adaptability to evolving optimization strategies, and optimized data handling with structured storage of results and logs. The LLM also assists in interpreting optimization outcomes in a user-friendly manner, offering advantages over traditional BI dashboards by providing increased flexibility, efficiency, and user accessibility in decision optimization. This LLM integration bridges the gap between advanced analytical models and practical applications, enabling more informed strategic planning and decision-making.

SE17

Building A Level 4 A403

Last Mile Planning and Operations

Invited Session

Practice Track

Chair: Chengliang Zhang, Amazon, Bellevue, United States

1 - Labor Capacity Planning in Last-Mile Operations

Chengliang Zhang, Amazon, Bellevue, WA, United States, Luying Sun

In the rapidly evolving landscape of last-mile delivery operations, labor capacity management is critical for operational excellence and customer satisfaction. With increased customer expectations and evolving business complexity across multiple business lines and regions, traditional workforce planning approaches are no longer sufficient. To adapt to these complex challenges, we present a mathematical programming model that simultaneously optimizes across multiple dimensions—including varied labor skills, flexible working schedules, and diverse delivery time windows. The model generates labor planning solutions that meet delivery commitments, satisfy business tenets, and optimize operational costs. The scalable framework effectively addresses the dynamic challenges of modern last-mile delivery operations while providing the flexibility needed for future growth and adaptation.

2 - Robotic Solution to Improve Transportation Utilization

Zhikun Gao, Amazon, Bellevue, WA, United States, Andrew Bruce, Jochen Koenemann

Containerization is introduced in the Amazon Fulfillment Network to provide an efficient cross-docking operation to ship customer orders between Amazon facilities. A drawback of currently leveraged container is that it leaves unused head-space in trailers. A stackable container is introduced to utilize more vertical space in trailers. A small scale pilot was launched in the US network, including fulfillment centers, sort centers, and delivery stations, to test the new container operational safety and assess key performance metrics. We conduct simulation to estimate the space utilization entitlement of the new container and design rigorous statistical experiments to measure operation benefits as well as risks. Also, we discuss preparations for extending the application of new container to a broader scope, such as regional or network-wide applications.

3 - Sequential Testing for Operations Experiments and Resource Ramp up

Brayan Ortiz, Amazon, Seattle, WA, United States

Supply chain operations frequently test new hardware, software and processes to improve productivity and quality, while avoiding downstream disruptions. In this talk, we provide an overview of sequential testing strategies for operational improvements, including classical methods for ramp-up and superiority detection. We explore how sequential testing approaches can be applied to various operational contexts, including buffered systems where traditional A/B testing may be less effective. Using real operations experiments, we compare traditional A/B testing with sequential methods, demonstrating significant program-level time savings across different operational scenarios. Our findings show that with sequential testing, we observed a 2-4 week reduction in testing time, and speeding up decision-making. This is particularly important to refine and continually improve innovations. We conclude with practical recommendations for implementing sequential testing methodologies in diverse operational environments, highlighting the potential for improved agility and responsiveness in process optimization efforts.

4 - Closing the Gap Between Scheduling and Execution of Under the Roof Processes

Nikolaos Lappas, amazon.com, Bellevue, WA, United States

We present a framework designed to enhance deterministic optimization models' ability to handle uncertainty through learned parameter offsets. Instead of relying on stochastic or robust optimization techniques, our framework offers a practical approach by systematically searching for optimal parameter adjustments using historical data. We present a scheduling case study, demonstrating how these offsets can effectively hedge against uncertainty while preserving the simplicity and interpretability of deterministic models. The framework's adaptability to improved parameter estimates and its support for both grid search and Bayesian experimental design make it a flexible tool for real-world applications. While our focus is on scheduling problems, the methodology is generalizable to other optimization domains where uncertainty plays a significant role. This work contributes to bridging the gap between theoretical optimization models and practical implementation in uncertain environments.

SE18

Building A Level 4 A404

Drones, UAVs, and Emergency Logistics

Invited Session

TSL: Intelligent Transportation Systems

Chair: Gino Lim, University of Houston, Houston, TX, United States

1 - A Dynamic Unmanned Aerial Vehicle Routing Framework for Urban Traffic Monitoring

Yumeng Bai, Purdue University, West Lafayette, IN, United States, Yiheng Feng

Unmanned Aerial Vehicles (UAVs) have great potential in urban traffic monitoring due to their rapid speed, cost-effectiveness, and extensive field-of-view, while being unconstrained by traffic congestion. However, their limited flight duration presents critical challenges in sustainable recharging strategies and efficient route planning for long-term monitoring. Additionally, existing approaches for long-term monitoring often neglect the dynamic and evolving nature of urban traffic networks. In this study, we introduce a novel dynamic UAV routing framework for long-term, network-wide traffic monitoring, leveraging existing ground vehicles (GVs), such as public transit buses, as mobile charging stations without disrupting their operations. To address the complexity of long-term monitoring involving multiple flights, we decompose the problem into manageable single-flight tasks, each modeled as a Team Arc Orienteering Problem with Decreasing Profits (TAOP-DP), aiming to collectively maximize the spatiotemporal network coverage. Between flights, edge weights are adaptively updated to incorporate real-time traffic changes and revisit intervals. We validate our framework through microscopic simulations on a modified Sioux Falls network under various scenarios, including static and changing demand levels and varying historical information availability. Comparative results demonstrate that our model outperforms three baselines, especially when historical information is incomplete or absent. Moreover, our monitoring framework captures network-wide traffic trends and constructs accurate Macroscopic Fundamental Diagrams (MFDs). These findings demonstrate the effectiveness of the proposed UAV routing framework, underscoring its suitability for efficient, reliable long-term traffic monitoring. The framework's adaptability and high accuracy in capturing the MFD highlight its potential for real-time network-wide traffic control management applications.

2 - Mothership-Based Drone Routing with TIME Windows via a Branch-and-Price-and-Cut Algorithm

Reza Mirjalili, University of Houston, Houston, TX, United States, Gino Lim, Jonathan Bard

Drones are transforming logistics, particularly in hybrid delivery systems where a truck acts as a mobile base—referred to as the *mothership*—from which drones are dispatched. This research addresses the Mothership-Based Drone Routing on a Network with Time Windows (MDRNTW) problem, in which the truck follows a fixed route (denoted as PPP), while drones perform deliveries within specified time windows. To address scalability issues inherent in arc-based formulations, we propose a path-based model using Dantzig-Wolfe reformulation and solve it via an exact Branch-and-Price-and-Cut (BPC) algorithm. The BPC framework decomposes the problem into a master problem (MP) and multiple subproblems (SPs). Although each SP can be solved efficiently using bidirectional label-setting, the number of SPs per iteration is large. Hence, we introduce an inferred bounding strategy based on previously solved SPs to skip those unlikely to yield columns with negative reduced cost, reducing column generation time by 32%. To strengthen the LP relaxation of the MP, we incorporate two valid inequalities and one novel cut, which resulted in an improved root node optimality gap.

3 - Optimizing Charging and Dispatch Decisions for Medical Drone Deliveries Under Uncertainty via Deep Reinforcement Learning Methods

Amin Asadi, University of Twente, Enschede, Netherlands, Abolfazl Maleki

Drones are increasingly employed in healthcare logistics to deliver medical supplies to remote, congested, or disaster-affected areas where traditional transportation systems are inadequate. However, efficiently managing the drone delivery system's operations remains challenging due to uncertain demand, limited fleet availability, and trade-offs associated with battery charging modes. Fast charging enhances short-term drone availability but accelerates battery degradation, reducing flight range and increasing replacement costs in the long term, while slow charging preserves battery life but limits immediate responsiveness.

We introduce SMART (Stochastic Multi-class Allocation and Recharging model for Transportation), a sequential decision-making framework that jointly optimizes drone dispatching and battery-recharging policies under resource- and operational-constraints. SMART explicitly considers heterogeneous flight-range requirements, stochastic spatiotemporal demand, and dynamic trade-offs between fast and slow charging. We formulate the system as a Markov Decision Process (MDP) that tracks battery charge levels, mission durations, and fleet availability over time. To derive real-time, adaptive policies, we employ multiple deep reinforcement learning agents based on DQN and DDQN. Computational experiments demonstrate that our algorithms for solving SMART significantly improve cost-efficiency, drone availability, and responsiveness to demand compared to traditional heuristic strategies. The resulting policies offer actionable insights into managing medical drone delivery systems through adaptive fast and slow charging strategies that ensure immediate responsiveness while minimizing long-term replacement costs and battery waste.

4 - From Auction to Delivery: Fast Resource Exchange and Collaborative Routing with Microswap

Ozge Aygul, Worcester Polytechnic Institute, Worcester, MA, United States, Yunus Doğan Telliel, Sarah Stanlick, Jen Pazour, Andrew Trapp

Nonprofit organizations (NPOs) provide essential resources and services in response to time-sensitive community needs. Yet, these NPOs often face challenges in quickly securing resources and coordinating with other organizations. To address these challenges, we introduce a novel framework to share and transport resources, designed for NPOs operating under a strict time horizon. Extending the SWAP platform for resource sharing (Huang et al., 2023), the MicroSWAP auction mechanism allows time-sensitive nonprofits to exchange resources by declaring their needs and available resources, while simultaneously providing optimized allocation and routing plans. In particular, upon reaching a certain market saturation, our framework identifies optimal resource exchanges and assigns and schedules pickup and delivery tasks through selective vehicle routing. Our approach leverages horizontal vehicle routing collaboration to optimize routes, improving service speed, efficient coordination among nonprofits, and collective impact. We demonstrate our model's scalability and effectiveness through real-world inspired case studies, highlighting its ability to mobilize collaboration in the nonprofit sector for time-sensitive community needs.

5 - Anticipating Infeasibility: Decision Diagram-Guided Reinforcement Learning for Tsp with Time Windows

Macarena Navarro, Carnegie Mellon University, Pittsburgh, PA, United States, Karan Singh, Willem-Jan van Hoeve, Segev Wasserkrug

Recent advances in neural combinatorial optimization have shown that deep reinforcement learning can solve combinatorial optimization problems by training policies to generate near-optimal solutions. However, existing approaches struggle with constraints that require global reasoning on the entire decision space, such as time window constraints augmenting the traveling salesperson problem. Most existing approaches only handle simpler constraints, such as permutation structures, where one can ensure feasibility incrementally while generating the solution.

We address this limitation by integrating relaxed decision diagrams into the reinforcement learning framework. Our approach equips the learner with structured lookahead capabilities that enable it to anticipate and avoid decisions leading to downstream infeasibility during solution generation. We embed these decision diagrams directly into attention-based neural architectures and evaluate their effectiveness on the traveling salesperson with time windows problem.

SE19

Building A Level 4 A405

Recent Advance on Interaction Between Machine Learning and Differential Equation

Invited Session

Applied Probability Society

Chair: Yiping Lu, IL, United States

1 - Optimal-PhiBE: a Pde-Based Model-Free Framework for Continuous-TIME Reinforcement Learning

Yuhua Zhu, University of California, Los Angeles, Los Angeles, CA, United States

This talk addresses continuous-time reinforcement learning (RL) in settings where the system dynamics are governed by a stochastic differential equation but remains unknown, with only discrete-time observations available. While the optimal Bellman equation enables model-free algorithms, its discretization error is significant when the reward function oscillates. Conversely, model-based PDE approaches offer better accuracy but suffer from non-identifiable inverse problems. To bridge this gap, we introduce Optimal-PhiBE, an equation that integrates discrete-time information into a PDE, combining the strengths of both RL and PDE formulations. Compared to the RL formulation, Optimal-PhiBE is less sensitive to reward oscillations, leading to smaller discretization errors. In linear-quadratic control, Optimal-PhiBE can even achieve accurate continuous-time optimal policy with only discrete-time information. Compared to the PDE formulation, it skips the identification of the dynamics and enables model-free algorithm derivation. Furthermore, we extend Optimal-PhiBE to higher orders, providing increasingly accurate approximations.

2 - Two Tales, One Resolution: Physics-Informed Test TIME Scaling and Precondition

Yiping Lu, Northwestern University, Evanston, IL, United States

In this talk, I will introduce a novel framework for physics-informed debiasing of machine learning estimators, which we call Simulation-Calibrated Scientific Machine Learning (SCaSML). This approach leverages the structure of physical models to achieve two key objectives:

Unbiased Predictions: It produces unbiased predictions even when the underlying machine learning predictor is biased.

Overcoming Dimensionality Challenges: It mitigates the curse of dimensionality that often affects high-dimensional estimators.

The SCaSML paradigm integrates a (potentially) biased machine learning algorithm with a de-biasing procedure that is rigorously designed using numerical analysis and stochastic simulation. Our methodology aligns with recent advances in inference-time computation—similar to those seen in the large language model literature—demonstrating that additional computation can enhance ML estimates.

Furthermore, we establish a surprising equivalence between our framework and another research direction that utilizes approximate (linearized) solvers to precondition iterative methods. This connection not only bridges two distinct areas of study but also offers new insights into improving estimation accuracy in complex, high-dimensional (PDE) settings.

3 - On the BSde-Based Methods for Solving Stochastic Optimal Control Problems

Baris Ata, University of Chicago, Booth School of Business, Chicago, IL, United States

I will give an overview of the BSDE-based methods for solving stochastic optimal control problems arising in operations management and provide a few applications from my recent work: Dynamic vehicle routing, dynamic bipartite matching, call-center scheduling, and stochastic joint replenishment problem. I will conclude with a discussion of effectiveness of the method.

4 - No-Regret Generative Modeling via Parabolic Monge-Ampère Pde

Nabarun Deb, University of Chicago, CHICAGO, IL, United States

We introduce a novel generative modeling framework based on a discretized parabolic Monge-Ampère PDE, which emerges as a continuous limit of the Sinkhorn algorithm commonly used in optimal transport. Our method performs iterative refinement in the space of Brenier maps using a mirror gradient descent step. We establish theoretical guarantees for generative modeling through the lens of no-regret analysis, demonstrating that the iterates converge to the optimal Brenier map under a variety of step-size schedules. As a technical contribution, we derive a new Evolution Variational Inequality tailored to the parabolic Monge-Ampère PDE, connecting geometry, transportation cost, and regret. Our framework accommodates non-log-concave target distributions, constructs an optimal sampling process via the Brenier map, and integrates favorable learning techniques from generative adversarial networks and score-based diffusion models. As direct applications, we illustrate how our theory paves new pathways for generative modeling and variational inference.

5 - Uniform-in-N Log-Sobolev Inequality for the Mean-Field Langevin Dynamics with Convex Energy

Matthew Zhang, UofT, Toronto, ON, Canada

We establish a log-Sobolev inequality for the stationary distribution of mean-field Langevin dynamics with a constant that is independent of the number of particles N . Our proof proceeds by establishing the existence of a Lipschitz transport map from the standard Gaussian measure via the reverse heat flow of Kim and Milman.

SE20

Building A Level 4 A406

RAS Student Paper Competition

Award Session

Railway Applications

Chair: Nikola Markovic, University of Utah, Salt Lake City, UT, United States

Co-Chair: Marcella Samà, Roma Tre University, Roma, Italy

Co-Chair: Pengling Wang, Tongji University, Shanghai

SE21

Building A Level 4 A407

Pierskalla Finalists Award Session

Award Session

Health Applications Society

Co-Chair: Jackie Baek, New York University, New York City, NY, United States

1 - Helping the Captive Audience: Advance Notice of Diagnostic Service for Hospital Inpatients

Miao Bai, University of Connecticut, Storrs, CT, United States, Nan Liu, Zheng Zhang

Inpatients are often treated as the “captive audience” for hospital diagnostic services, and they are notified only when service capacity is available. This arrangement causes significant chaos and inefficiencies in hospital operations. We propose an innovative scheduling approach called “advance notice” to manage hospital diagnostic practices. Advance notice is a brand-new scheduling paradigm in between classic allocation scheduling and advance scheduling. Patients are placed in a common queue waiting to be called for service, and they will be provided both a fixed preparation time and a guaranteed service time window in advance. The advance notice policy enjoys the benefit of allocation scheduling (giving the provider flexibility in using her capacity) and that of advance scheduling (reducing uncertainties in patient waiting for service). It calls for two decisions: who to serve now and who to send advance notices to. We formulate a Markov decision process model to optimize these decisions dynamically. Via a novel variable transformation, we reveal the hidden anti-multimodular structure of the problem and show how the optimal decisions should be adjusted in response to changes in the system load. Beyond solving the model for daily operations, we further investigate how the service and preparation time windows can be used as strategic levers to manage system performance. Our numerical study, populated by real data from a large academic medical center in the United States, demonstrates significant improvement in operational efficiency by switching from current practice to our proposed advance notice policy.

2 - Reducing Overdose Deaths and Mitigating County Disparities: Optimal Allocation of Substance use Treatment Centers

Matt Baucum, Oklahoma State University, Stillwater, OK, United States, Matt Harris, Larry Kessler, Guanyi Lu

Substance use disorder (SUD) is a pressing health concern in the U.S., and connecting communities with treatment is a growing public health imperative. We study the optimal allocation of substance use treatment centers across counties within U.S. states over a multi-year time horizon. We use a three-part objective function which minimizes (1) statewide overdose deaths, (2) mismatch between counties' treatment center access and overdose death rates ('inequity'), and (3) mismatch between counties' treatment center access and population ('inequality'). We also consider the temporal interdependence of counties' treatment center allocations and their death-rate-based 'equity' targets. We use a predict-then-optimize approach based on data from over 2,500 U.S. counties between 1999 and 2016. We first estimate SUD treatment centers' causal impact on county-level overdose death rates, using mental health parity laws as an instrument. We then use these estimates in a nonlinear program that reallocates states' observed levels of SUD treatment funding. Results suggest that this approach can meaningfully reduce statewide overdose deaths and between-county disparities in treatment access. Prioritizing death rate reductions over equity/equality results in more resources being directed to high-population counties, where treatment centers are most effective. Prioritizing equity/equality in treatment access results in more resources being directed to under-served low-population counties. We also find that states differ in their Pareto-optimal tradeoffs between death rates and inequity/inequality.

3 - Improving Access to Essential Medicines in Sierra Leone via Decision-Aware Machine Learning

Tsai-Hsuan (Angel) Chung, The Wharton School, UPenn, Philadelphia, PA, United States, Jatu Abdulai, Patrick Bayoh, Lawrence Sandi, Francis Smart, Hamsa Bastani, Osbert Bastani

A critical challenge in healthcare systems in Low- and Middle-Income Countries (LMICs) is the efficient and equitable allocation of scarce resources, particularly essential medicines. This problem is complicated by limited high-quality data, which restricts the applicability of traditional data-driven techniques. We propose a novel machine learning framework for essential medicines allocation, which leverages a combination of multi-task learning and decision-aware learning to improve sample efficiency and ensure equitable allocation. In collaboration with the Sierra Leone national government, our framework has been deployed nationwide as a decision support tool to help reduce waste and improve essential medicines allocation. Our evaluation using synthetic difference-in-differences analysis demonstrates a 19% increase in medicine consumption, with no changes to the supply, improving access for approximately 3.7 million women and children under five. Through experimental validation, we demonstrate that our approach also significantly outperforms baseline approaches. Our work demonstrates the tangible impact of machine learning in optimizing high-stakes decisions in resource-constrained settings, improving efficiency while ensuring equity and cost-effectiveness.

4 - Optimizing Initial Screening for Colorectal Cancer Detection with Adherence Behavior

Sarah Yini Gao, Singapore Management University, Singapore, Singapore, Yan He, Rujie Zhang, Zhichao Zheng, Sean Lam Shao Wei, Emile Tan

Two-stage colorectal cancer (CRC) screening programs typically begin with a fecal hemoglobin (f-Hb) test, where positive results lead to a follow-up colonoscopy. However, not all individuals adhere to the recommended second-stage test. We study how the design of the first-stage test—specifically, the choice of f-Hb cutoff values—affects both detection effectiveness (CRC and polyp identification) and efficiency (colonoscopy demand and cost). We develop an information design model that combines Bayesian persuasion with information avoidance behavior to analyze individual decision-making. The model is calibrated using diverse data sources, including a nationwide survey of 3,920 respondents in Singapore. We find that using a single cutoff can improve follow-up adherence, while revealing exact f-Hb values enhances detection. Raising the cutoff from the current standard to 39 µg/g could detect 21% more CRC and polyp cases, reduce unnecessary colonoscopies by 27%, and lower lifetime CRC risk by 11%. This policy change could also result in estimated cost savings of \$20 million in public healthcare spending and \$12 million in individual screening costs. Optimizing the cutoff for first-stage screening tests is key to balancing public health outcomes and resource use. The prevalent practice of using low thresholds to maximize sensitivity may be counterproductive, increasing false positives and straining follow-up adherence. Our findings provide evidence-based guidance for more effective CRC screening policy design.

SE22

Building A Level 4 A408

Energy and Applied Probability

Invited Session

Applied Probability Society

Chair: Gal Mendelson, Technion, Haifa, Israel

1 - Optimization Under Rare Event Constraints

Bert Zwart, CWI Amsterdam and TU Eindhoven, Amsterdam, Netherlands

We consider a class of chance-constrained programs in which profit needs to be maximized while enforcing that a given adverse event remains rare. Using techniques from large deviations and extreme-value theory, we show how the optimal value scales as the prescribed bound on the violation probability becomes small and how convex programs emerge in the limit. We also use our results to show that the popular CVaR approximation is asymptotically optimal under light-tail assumptions, but it is sub-optimal in a heavy-tailed setting. Then, we apply point process techniques and random set theory to study the suboptimality gap in the Monte Carlo based scenario approach.

2 - Failure-Aware Dynamic Reserve and Checkpoint Scheduling for Sustainable Gpu Clusters

Weici Pan, Stony Brook University, Stony Brook, NY, United States, Tina Xing, Zhenhua Liu

Increasing demands of large-scale AI training and other workloads on GPU clusters pose both environmental and economic challenges due to stochastic GPU failures, volatile electricity prices, and fluctuating carbon intensities. We present a probabilistic framework that co-optimizes operational cost and CO2 emissions under realistic GPU failure behavior and time-varying resource characteristics. Our model captures dynamic reserve allocation (spare GPUs) and checkpoint scheduling as jointly controlled decision variables, integrating real-world electricity price and carbon intensity traces alongside cluster energy-efficiency parameters in each time slot.

We formulate a weighted objective that combines expected monetary expense and emissions, governed by a tunable trade-off parameter. In the joint decision problem as a stochastic optimization over a finite horizon, at each epoch, the system chooses a reserve level to hedge against failures and a checkpoint interval to bound rollback losses. GPU failures are represented by heavy-tailed distributions calibrated on production cluster logs. Reserve decisions trade off idle-holding and provisioning costs against failure-count probabilities. While checkpoint intervals are selected to minimize the sum of I/O overhead and expected rollback penalty based on the estimated mean time between failures

In simulation studies driven by real price, carbon, and failure traces, our dynamic policy consistently outperforms static reserve and checkpoint baselines on the combined cost-emissions metric across a range of failure regimes and cluster scales. These results demonstrate a practical pathway to embedding failure awareness and environmental objectives directly within GPU cluster fault-tolerance protocols.

3 - Learning Decision-Focused Uncertainty Representations for Robust and Risk-Constrained Optimization

Christopher Yeh, California Institute of Technology (Caltech), Pasadena, CA, United States, Nicolas Christianson, Adam Wierman, Yisong Yue

While deep learning models often achieve high predictive accuracy, their predictions typically do not come with any provable guarantees on risk or reliability, which are critical for deployment in high-stakes applications, such as medical diagnoses or controlling the power grid. Recent tools from conformal inference offer distribution-free, finite-sample guarantees for uncertainty calibration, but such tools generally have not been applied to controlling tail risks, and they are generally agnostic to downstream decisions, thus leaving significant room for performance improvement. In this work, we develop an end-to-end framework for learning decision-focused uncertainty representations that provide rigorous guarantees while optimizing downstream decision performance. We instantiate our framework in two contexts: robust optimization, and CVaR-constrained optimization. For robust optimization, our approach combines conformal prediction with partially input-convex neural networks to approximate arbitrary convex uncertainty sets. For CVaR-constrained optimization, we extend the framework of conformal risk control (CRC) to handle a broad class of Optimized Certainty Equivalent (OCE) risks, including conditional value-at-risk (CVaR). To overcome the limitations of post-hoc calibration, we introduce Conformal Risk Training, which differentiates through the conformal procedure to jointly learn model parameters and the uncertainty representation. Across applications in energy storage arbitrage, portfolio optimization, and medical image classification, our framework consistently outperforms standard two-stage or post-hoc approaches—achieving both provable robustness and improved average-case performance.

4 - Stochastic Optimization and Scenario Generation for Electric Grid Resilience

John Hasenbein, University of Texas at Austin, Austin, TX, United States, Ashutosh Shukla, Shunan Zheng

In this work we develop models to help decision makers increase resilience in electric grids when facing extreme weather events such as floods and heat waves. Our case studies involve the electric grid in Texas, which frequently suffers from hurricane and other flood events. A major challenge is to develop a sufficiently representative collection of scenarios in order to populate large, long-term, stochastic optimization models. We discuss classical methods such as NORTA, and more modern ML-style approaches to this problem.

SE23

Building A Level 4 A409

Session Title: Academic and Industry Job Search

Panel Session

Minority Issues Forum

Co-Chair: Carmen Haseltine, University of Wisconsin-Madison, Madison, WI, United States

1 - Panelist

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

2 - Panelist

Emily Tucker, Clemson University, Clemson, SC, United States

3 - Panelist

Thiago Serra, University of Iowa, Iowa City, IA, United States

4 - Panelist

Carmen Haseltine, University of Wisconsin-Madison, Madison, WI, United States

SE24

Building A Level 4 A410

Responsible Machine Learning via Mixed-Integer Optimization

TutORial session

TutORials

Chair: Nilay Noyan, AMAZON, Bellevue, WA, United States

1 - Responsible Machine Learning via Mixed-Integer Optimization

Nathan Justin, University of Southern California, Los Angeles, CA, United States, Qingshi Sun, Andres Gomez, Phebe Vayanos

In the last few decades, machine learning (ML) has achieved significant success across domains ranging from healthcare, sustainability, and the social sciences to criminal justice and finance. But its deployment in increasingly sophisticated, critical, and sensitive areas affecting individuals, the groups they belong to, and society as a whole raises critical concerns around fairness, transparency, and robustness, among others. As the complexity and scale of ML systems and of the settings in which they are deployed grow, so does the need for responsible ML methods that address these challenges while providing guaranteed performance in deployment. Mixed-integer optimization (MIO) offers a powerful framework for embedding responsible ML considerations directly into the learning process while maintaining performance. For

example, it enables learning of inherently transparent models that can conveniently incorporate fairness or other domain-specific constraints. This tutorial paper provides an accessible and comprehensive introduction to this topic discussing both theoretical and practical aspects. It outlines some of the core principles of responsible ML, their importance in applications, and the practical utility of MIO for building ML models that align with these principles. Through examples and mathematical formulations, it illustrates practical strategies and available tools for efficiently solving MIO problems for responsible ML. It concludes with a discussion on current limitations and open research questions, providing suggestions for future work.

SE28

Building B Level 2 B202

Optimization Under Uncertainty in Service Systems

Invited Session

Computing Society

Chair: Aliaa Alnaggar, Toronto Metropolitan University, Toronto, ON, Canada

Co-Chair: Mucahit Cevik, Toronto Metropolitan University, Toronto, ON, Canada

1 - Sample-Path Clustering for Context-Sensitive Markov Decision Processes

Allison Grimsted, University of Michigan, Ann Arbor, MI, United States, Agni Orfanoudaki, Brian Denton

We address the problem of estimating transition probability matrices for heterogeneous longitudinal data representing sample-paths for stochastic systems. We formulate the clustering as a set partitioning problem and use column generation with a mixed integer nonlinear pricing problem to assign sample-paths to a parsimonious set of clusters, each associated with its own transition probability matrix to capture contextual variation in the system state transition probabilities. We benchmark our approach against a baseline maximum likelihood estimation method that pools all sample-paths to estimate a single "one-size-fits-all" transition matrix, using a highway bridge maintenance case study.

2 - Convexification Methods for Solving Two-Stage Distributionally Robust Optimization with Applications to Crowdsourced Delivery

Aliaa Alnaggar, Toronto Metropolitan University, Toronto, ON, Canada, Merve Bodur

We consider a service region planning problem faced by a crowdsourced delivery platform where drivers are commuters who are willing to deviate from their original routes to a make a delivery in exchange for a compensation. The availability of drivers and service demand are uncertain, and it is difficult to estimate their probability distributions due to the absence of service history. To mitigate the effects of data ambiguity, we propose a two-stage distributionally robust optimization (DRO) model. The first stage selects which nodes to offer service in, while the second stage matches drivers and orders after uncertainty is realized. We derive an exact reformulation of the DRO model and develop a monolithic approximation based on a convex relaxation of the subproblem. We further strengthen the proposed approximation by a set of valid inequalities inspired by the linearization reformulation technique. The benefit of the proposed approach for service region design in crowdsourced delivery is demonstrated through numerical experiments.

3 - Integrated Dynamic Pricing and Dispatching for Crowd-Shipping via Neural Approximate Dynamic Programming

Mucahit Cevik, Toronto Metropolitan University, Toronto, ON, Canada, Arash Dehghan, Merve Bodur

Crowd-shipping, which utilizes non-professional couriers for deliveries, presents a scalable and sustainable alternative to traditional logistics models. This work introduces a Neural Approximate Dynamic Programming (NeurADP) framework, which combines ADP and deep reinforcement learning (DRL), to optimize crowd-shipping operations. More specifically, we integrate dynamic pricing and dispatching to enhance efficiency and profitability. Extending prior work, our model incorporates features such as batching and courier queues, addressing the complexities of large-scale operations. Our experiments comparing NeurADP policy with myopic and DRL baselines demonstrate significant improvements in operational efficiency, cost reduction, and sustainability. Sensitivity analyses highlight the impact of pricing strategies, courier availability, and delivery constraints, offering actionable insights for real-world applications.

4 - Optimal Catheter Replacement Planning for Newborn Care

F. Safa Erenay, University of Waterloo, Waterloo, ON, Canada, Atahan Bayir, Cansu Dagsuyu, Sibel Alumur

Catheters facilitate fluid delivery and waste removal for vulnerable newborns. However, their need for frequent replacement increase the risk of infections depending on catheter type, and replacement time and sequence. The infection risk can be minimized by optimizing catheter replacement decisions. For this purpose, we develop a Markov Decision Process (MDP) model incorporating generally distributed stochastic hospital stay and catheter life time using non-homogeneous Bernoulli trials. The model dynamically tracks catheterization history in the state space, and performance metrics such as infection rate, cost, catheter lifetime wastage to balance infection control and operational efficiency. The proposed approach can reduce the infection risks by up to five-fold, and costs and material waste by two- to three-fold.

5 - Stochastic Revelation in Sequential Decision-Making with a Pharmaceutical Application

morteza davari, SKEMA Business School, Lille, France, Merve Bodur, Bernardo Pagnoncelli

In many real-life situations, such as medical product launches, energy investments, or the rollout of new policies, decision-makers must act before knowing exactly when critical information will become available. We develop new mathematical models that incorporate uncertainty about what will happen and when that uncertainty will resolve. Traditional decision-making tools assume fixed timelines for information revelation; instead, we address more realistic scenarios where information arrives at unpredictable moments, making planning more complex and costly if mishandled.

In this talk, we focus on a pharmaceutical application, in particular, the planning and production of medical devices, where firms must be mindful of regulatory approvals. We develop a framework enhanced by problem-specific decomposition methods to address this multistage stochastic problem.

SE29

Building B Level 2 B203

Recent Advances in Deep Learning and Quantum Computing for AC Optimal Power Flow

Invited Session

ENRE: Energy

Chair: Yu Zhang, UC Santa Cruz, Santa Cruz, CA, 95064, United States

1 - Quantum Computing for Solving the AC-OPF

Vassilis Kekatos, Purdue University, West Lafayette, IN, United States

This work puts forth a hybrid quantum-classical approach to cope with large-scale instances of the AC optimal power flow (OPF) problem. We advocate encoding primal and dual variables as states of variational quantum circuits (VQC) and pursuing saddle points of the associated Lagrangian function via primal-dual algorithms. VQCs are parameterized quantum circuits, which can be trained using classical optimization methods and are well-suited for near-term quantum computers. Leveraging the power system structure, we also put forth novel protocols for efficiently measuring the quantum observables associated with power system optimization. Preliminary numerical tests on quantum simulators study the complexity-performance trade-off of the suggested methods.

2 - Unsupervised Learning for Solving AC Optimal Power Flows: Design, Analysis, and Experiment

Minghua Chen, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China, People's Republic of, Wangjun Huang, Steven Low

With the increasing penetration of renewables, AC optimal power flow (AC-OPF) problems need to be solved more frequently for reliable and economic power system operation. Supervised learning approaches have been developed to solve AC-OPF problems fast and accurately. However, due to the non-convexity of AC-OPF problems, it is non-trivial and computationally expensive to prepare a large training dataset, and multiple load-solution mappings may exist to impair learning even if the dataset is available. In this work, we develop an unsupervised learning approach (DeepOPF-NGT) that does not require ground truths. DeepOPF-NGT utilizes a properly designed loss function to guide neural networks in directly learning a legitimate load-solution mapping. Kron reduction is used to remove the zero-injection buses from the prediction. To tackle the unbalanced gradient pathologies known to deteriorate the learning performance, we develop an adaptive learning rate algorithm to dynamically balance the gradient contributions from different loss terms during training. Further, we derive conditions for unsupervised learning to learn a legitimate load-solution mapping and avoid the multiple mapping issue in supervised learning. Results of the 39/118/300/1354-bus systems show that DeepOPFNGT achieves optimality, feasibility, and speedup performance comparable to the state-of-the-art supervised approaches, and a few ground truths can further improve its performance. This is a joint work with Wanjun Huang and Steven Low.

3 - Accelerated Learning Framework for AC Optimal Power Flow

Yu Zhang, UC Santa Cruz, Santa Cruz, CA, United States

We propose a physics-informed, semi-supervised learning framework for real-time AC optimal power flow (AC-OPF). To reduce dependence on slow solvers, we use ridge regression to generate pseudo-labels and augment training data efficiently. We introduce novel gradient estimation techniques—including batch-mean, linearized, and decoupled Jacobians—and a reduced branch set method to accelerate training while ensuring constraint satisfaction. Tested on large-scale systems, our approach achieves near-optimal, feasible solutions with over 30× speedup in inference and 10× faster training, making it practical for real-time grid operations.

4 - Neural Approximation for Distribution Optimization

Mohamad Chehade, University of Texas at Austin, Austin, TX, United States

The rise of distributed energy resources (DERs) is reshaping modern distribution grids, introducing new challenges in attaining voltage stability under dynamic and decentralized operating conditions. This paper presents NEO-Grid, a unified learning-based framework for volt-var optimization (VVO) and volt-var control (VVC) that leverages neural network surrogates for power flow and deep equilibrium models (DEQs) for closed-loop control. Our method replaces traditional linear approximations with piecewise-linear ReLU networks trained to capture the nonlinear relationship between power injections and voltage magnitudes. For control, we model the recursive interaction between voltage and inverter response using DEQs, allowing direct fixed-point computation and efficient training via implicit differentiation. We evaluated NEO-Grid on the IEEE 33-bus system, demonstrating that it significantly improves voltage regulation performance compared to standard linear and heuristic baselines in both optimization and control settings. Our results establish NEO-Grid as a scalable, accurate, and interpretable solution for learning-based voltage regulation in distribution grids.

SE30

Building B Level 2 B204

Game Theory Applications in Power Systems

Invited Session

ENRE: Electricity

Chair: Wei Wang, Pacific Northwest National Laboratory, Richland, WA, United States

Co-Chair: Liping Li, Pacific Northwest National Laboratory, Columbus, OH, United States

1 - Gaming on Coincident Peak Shaving: Equilibrium and Strategic Behavior

Bolun Xu, Columbia University, New York, NY, United States, Liudong Chen, Jay Sethuraman

Power system operators and electric utility companies often impose a coincident peak demand charge on consumers when the aggregate system demand reaches its maximum. This charge incentivizes consumers to strategically shift their peak usage away from the system's collective peak, which helps reduce stress on electricity infrastructure. In this talk, we develop a game-theoretic model to analyze how such strategic behavior affects overall system efficiency. We show that depending on the extent of consumers' demand-shifting capabilities, the resulting coincident peak shaving game can exhibit concavity, quasi-concavity with discontinuities, or non-concavity with discontinuities. In a two-agent, two-period setting, we derive closed-form Nash equilibrium solutions for each scenario and generalize our findings to multi-agent contexts. We prove the stability of the equilibrium points and propose an algorithm for computing equilibrium outcomes under all game configurations. Our results indicate that while the decentralized game achieves peak-shaving outcomes comparable to those of a centralized model, it does so at the cost of higher anarchy. Under quasi-concave and non-concave conditions, this inefficiency grows with increased consumer flexibility and larger disparities in marginal shifting costs; we also examine how the number of agents influences system performance. Finally, numerical simulations validate our theoretical insights.

2 - A Stackelberg Game Model for Incentive Design to Align Distributed Energy Resource Investments with System Objective

Liping Li, Pacific Northwest National Laboratory, Richland, WA, United States, Wei Wang, Sarah Newman, Thiagarajan Ramachandran

Distributed Energy Resources (DERs) play a crucial role in enhancing the reliability and resilience of power systems at the community level. Their decentralized nature necessitates innovative strategies to promote coordinated and efficient investment. This presentation introduces a Stackelberg game theory model to create incentives that stimulate DER investments, thereby achieving system-wide objectives. In this model, a system operator acts as the leader, offering incentives to foster DER investments, such as photovoltaic panels and energy storage devices, within communities and microgrids. The community and microgrid operators, functioning as followers, aim to maximize their own benefits. By employing strategic incentive design, the leader influences the followers' investment and operational decisions, optimizing the overall welfare of the system. The presentation will include a demonstration and analysis of results from a case study involving three communities, focusing on minimizing the total unsatisfied demand across all communities.

3 - Design of Multi-Day Capacity Option Products to Improve Unit Compliance in Real-Time

qiwei zhang, University of Texas San Antonio, San Antonio, TX, United States

Generators need to secure fuel (e.g., natural gas contracts) in advance to ensure they can operate in real-time. This results in fuel advancement costs that are incurred before any revenue is realized from short-term market participation. However, current market settlements do not explicitly compensate these upfront costs. Consequently, suppliers face a disincentive to secure resources in advance, which may lead to insufficient availability during system needs. To address this gap, we propose a new payment mechanism that compensates generators for fuel advancement. Under this scheme, generators receive an upfront payment—a premium—for ensuring adequate resource is available ahead of operations. This payment reflects the value of fuel availability and covers expenses such as early fuel procurement, thereby enhancing system reliability.

4 - Toward Efficient Integration of Grid-Edge Flexibility in Voltage Regulation

Zhirui Liang, Johns Hopkins University, Baltimore, MD, United States, Rajni Bansal, Joshua Comden, Andrey Bernstein, Yury Dvorkin

This presentation addresses the challenges of integrating grid-edge Distributed Energy Resources (DERs) for voltage regulation in modern distribution systems. With increasing DER penetration, Distribution System Operators (DSOs) face issues such as uncertainty and limited visibility. DER Aggregators (DERAs) can mitigate these challenges by coordinating DERs for market participation and ancillary services, including voltage regulation. However, the role of DERAs in voltage regulation remains underexplored. We compare two approaches for procuring flexibility from DERAs: a market-based method, where the DSO and DERAs make decisions simultaneously through price signals, and an incentive-based method, where the DSO leads by offering incentives. Using game theory, we analyze the efficiency of each approach, examining equilibrium outcomes, social welfare, and the impact of deviations from ideal conditions. The study provides a detailed theoretical and numerical comparison, offering insights into which procurement method better supports voltage regulation. Our findings contribute to the design of more effective grid-edge management strategies, enhancing DSO-DERA coordination for voltage regulation.

5 - Learning in Stackelberg Games and Applications in Electricity Rate Design

Andrew Liu, Purdue University, West Lafayette, IN, United States, Jun He, Yihsu Chen

Designing socially optimal policies in multi-agent environments is a fundamental challenge in both economics and artificial intelligence. This paper studies a general framework for learning Stackelberg equilibria in dynamic and uncertain environments, where a single leader interacts with a population of adaptive followers. Motivated by pressing real-world challenges such as equitable electricity tariff design for consumers with distributed energy resources (such as rooftop solar and energy storage), we formalize a class of Stackelberg Markov games and establish the existence and uniqueness of stationary Stackelberg equilibria under mild continuity and monotonicity conditions. We then extend the framework to incorporate a continuum of agents via mean-field approximation, yielding a tractable Stackelberg–Mean Field Equilibrium (S-MFE) formulation. To address the computational intractability of exact best-response dynamics, we introduce a softmax-based approximation and rigorously bound its error relative to the true Stackelberg equilibrium. Our approach enables scalable and stable learning through policy iteration without requiring full knowledge of follower objectives. We validate the framework on an energy market simulation, where a public utility or a State energy commission sets time-varying rates for a heterogeneous population of prosumers. Our results demonstrate that learned policies can simultaneously achieve economic efficiency, equity across income groups, and stability in energy systems. This work demonstrates how game-theoretic learning frameworks can support data-driven policy design in large-scale strategic environments, with applications to real-world systems like energy markets.

SE31

Building B Level 2 B205

Managing Online Retailers and Services

Invited Session

eBusiness

Chair: Hongmin Li, Arizona State University, Tempe, AZ, United States

Co-Chair: Yuqi Yang, Arizona State University, Tempe, AZ, United States

1 - Goodrx Effect in the Pharmaceutical Value Chain

Yishen Cai, University of Miami Herbert Business School, Miami, FL, United States, Xin Geng, Hongmin Li, Nan Yang

In recent years, coupon vending has become a popular method for patients to bypass high usual and customary (UC) off-counter prescription drug prices. GoodRX is most successful for this practice---they provide steeply discounted drug prices (often exceeding 60% off), for almost no charge to anyone with or without insurance, on a digital coupon redeemable at most popular pharmacies. We uncover the enabling value chain mechanism and examine the GDRX effect focusing on profit reallocation in the pharmaceutical retail market.

2 - Asymptotic Optimality and Performance Bounds of Partially Personalized Fixed Pricing Policies

ZHUYU LIU, the University of British Columbia, Vancouver, BC, Canada, Meichun Lin, Woonghee Huh

We study a dynamic pricing problem involving a single product, finite inventory, and multiple customer types with stochastic demand over a finite horizon. The retailer observes the customer's type before offering a price and must determine whether to employ fully dynamic personalized pricing or simpler fixed pricing strategies that group customers into a limited number of segments. Motivated by operational and fairness constraints, we analyze the theoretical performance of fixed pricing policies under varying degrees of personalization relative to the fully personalized dynamic pricing benchmark.

We first derive an asymptotically tight lower bound for the fully personalized fixed-price (FP) policy by leveraging a deterministic relaxation of the stochastic problem. We show that in the high-volume regime where both inventory and planning horizon scale proportionally, this policy is asymptotically optimal, and we provide a convergence rate. We also study a class of K-personalized FP policies, where customer types are partitioned into K groups, each offered a fixed price. We establish lower bounds for such policies. Our results provide a theoretical framework for evaluating the revenue implications of personalization granularity under fixed pricing

3 - Return Rate Estimation Under Complete and Incomplete Data

Eleanor Wang, Cornell University, Scottsdale, AZ, United States, Yuqi Yang, Hongmin Li, Miao Zheng

Fast fashion retailers like Zara and Shein have generous, cost-free return policies which stimulate demand but also incur high reverse-logistics costs, highlighting the need to analyze product returns, which can then be utilized to make sales decisions.

Our paper develops a model for product returns – specifically, we address the challenge of estimating SKU-level return rates when only aggregate return counts are available. Motivated by an Amazon apparel seller, we develop return rate estimation methods under two cases. In the complete-data case, when SKU-level returns are observed, we derive a maximum-likelihood estimator to obtain unbiased return-rate estimates, and prove its asymptotic efficiency. In the case of incomplete data, when only total returns are available, we solve the issue of identifiability of the maximum likelihood under classical distributional approximation. Finally, we embed our return-rate estimates into a price-optimization model to demonstrate practical gains and to inform optimal pricing decisions for retailers.

4 - Learning for New Products Under Costly Experimentations

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

We consider an online retailing firm that leverages transaction-level data to guide sequential product experimentation. The firm is myopic, and new products are generated by recombining features of previously tested products. Each untested product is associated with a Bayesian posterior belief whose precision increases with its similarity to the experimented products.

We formulate this decision problem as a discrete-time, infinite-horizon dynamic search in which the state comprises the set of tested products and their realized performances. We establish the existence and uniqueness of the most promising product at each decision epoch. The firm then compares the expected improvement from experimenting on this candidate against the per-trial cost of delaying commitment. We prove that the optimal expected gain from search decreases in expectation over time. Our results yield tractable, threshold-based decision rules that can be readily implemented by online-retailing firms seeking to exploit rich product-performance data for innovative product development.

SE33

Building B Level 2 B207

Healthcare Policy Optimization

Invited Session

Health Applications Society

Chair: Hyunwoo Shin, Virginia Tech, Blacksburg, VA, United States

Co-Chair: Sait Tunc, Virginia Tech, Blacksburg, VA, United States

1 - A Framework to Incorporate Human Behavior in Agent-Based Models for Infectious Disease

Maria Mayorga, North Carolina State University, Raleigh, NC, United States, Sebastian Rodriguez-Cartes, Osman Ozaltin, Julie Swann

Computational agent-based model (ABM) simulations are critical for projecting and responding to infectious disease outbreaks, yet the effectiveness of health interventions heavily relies on individual adherence, underscoring the importance of accurately representing human behavior. While some ABMs have incorporated behaviors, clear guidelines for leveraging the extensive literature on behavioral theories have been lacking. This study addresses this challenge by proposing a modeling framework that integrates classical behavioral theories and their psychological constructs, drawing upon well-established definitions and categorizations from the social sciences to guide operationalization. We present novel and existing modeling approaches to represent these constructs and their interactions with the environment and behavior. Furthermore, we discuss methods for determining an agent's likelihood of adopting a behavior and how empirical data can be used for calibration and validation. To illustrate the framework's application to disease-specific behaviors, we provide a modeling example. Ultimately, this proposed framework offers a roadmap for computational modelers to incorporate scientifically grounded behaviors in the design and evaluation of simulation models and interventions.

2 - Design of a Non-Profit Telemedicine-to-Hospital Network for Specialized Healthcare Services

Yuting Yuan, Penn State, University Park, PA, United States, Lan Jin, Hui Zhao, Nicholas Petrucci

Reliable and equitable access to specialized public healthcare services is a challenge for society. We investigate the optimal design of a Telemedicine-to-Hospital (T2H) network for a non-profit center that provides such a service to participating hospitals. If a hospital chooses to participate in the center's network, then all patients visiting the hospital in need of the specialized service are treated remotely by the telemedicine center with coordination from the hospital. The center's design problem is to determine the two-part hospital fee policy and corresponding staffing level to maximize total patient welfare subject to service, cost, and hospital participation constraints. This problem is inspired by a telemedicine center that provides treatment for victims of sexual assault. We develop a two-stage Stackelberg game, first, to determine the center's optimal solution and, second, to analyze the nuanced implications for external government funding, which helps provide such services. Our calibrated numerical study based on real data further investigates how the size and mix of hospitals within the center's service reach affect its optimal design. To the best of our knowledge, this is among the first studies designed to address the salient trade-offs and financial challenges of a non-profit T2H telemedicine network for specialized healthcare services. Among other things, our research demonstrates how well-designed fees paired with efficient levels of government funding can help expand public access to high-quality specialized healthcare services. Our work develops a much-needed model framework to provide these expensive yet critical public healthcare services to society.

3 - Dynamic Capacity Planning to Improve Healthcare Access

Duygu Soylemez, The University of Chicago, Booth School of Business, Chicago, IL, United States, Daniel Adelman, Kenneth Nunes

We examine a stylized appointment planning problem motivated by access issues in healthcare. Our analysis reveals structural insights and identifies optimal policies under certain conditions.

4 - Diagnostics to Detect *Dracunculus Medinensis* Analytes in Biospecimens in the Last Mile of Disease Eradication: the Impacts of Diagnostic Accuracy, Deployment Protocol and Tethering Compliance

Hannah Smalley, ISyE Georgia Tech, Atlanta, GA, United States, Pinar Keskinocak, Maryann Delea, Obiora Eneanya, Philip Ouakou Tchindebet, Jessica Van Loben Sels, Adam Weiss

Guinea worm disease (GWD) has been targeted for eradication, and much progress has been made towards this goal, with a 99.99% reduction in human cases since 1986. However, animal infections, particularly among dogs, create challenges to reaching disease eradication. Diagnostics to detect *Dracunculus medinensis* analytes in dogs are under development with the goal of identifying infected dogs prior to worm emergence and thus providing national Guinea worm eradication programs with additional data to inform decisions about intervention implementation, such as proactive tethering of dogs in GW-endemic areas to prevent contamination of water sources with Guinea worm larvae and interrupt disease transmission. We use infection data provided by Chad's Guinea Worm Eradication Program in an agent-based simulation which models Guinea worm infections among dogs and interactions between dogs and water sources to determine the impact of interventions to prevent GWD over time. Based on the potential characteristics of diagnostics that could detect prepatent, recent, or past GW infections, we analyze a variety of testing accuracy levels (i.e., test sensitivity and specificity) and deployment protocols which incorporate screening and secondary diagnostics. Diagnostic testing can have a positive impact due to the increase in tethering triggered by dogs testing positive for GWD, particularly in areas having low compliance with proactive tethering recommendations. Increasing compliance with proactive tethering reduces the impact of diagnostic testing.

SE34

Building B Level 2 B208

Recent Techniques in Multi-Level Optimization

Invited Session

OPT: Global Optimization

Chair: Anil Aswani, UC Berkeley, Berkeley, CA, United States

Co-Chair: Sukanya Kudva, UC Berkeley, Berkeley, CA, United States

1 - Algorithm for Trilevel Optimization

Sukanya Kudva, UC Berkeley, Berkeley, CA, United States, Anil Aswani

Multi-level optimization problems can help the industry and the government in making decisions that involve a hierarchy of players or a chronology of events in time. The first player in a hierarchy or the first chronological event takes the first action, followed by the next in hierarchy/ chronology, and so on. And at each level, the action optimizes an objective subject to the actions previously taken and is dependent on future actions that could be taken. In this work, we propose an algorithm for a trilevel optimization problem, when there are three such levels. This algorithm can be used when actions are taken in a continuous space of variables, and we assume quadratic objectives and linear constraints. We then prove the convergence of our algorithm in a finite number of iterations.

2 - Mathematical Program Networks

Forrest Laine, Vanderbilt University, Nashville, TN, United States

Mathematical Program Networks (MPNs) are introduced in this work. An MPN is a collection of interdependent Mathematical Programs (MPs) which are to be solved simultaneously, while respecting the connectivity pattern of the network defining their relationships. The network structure of an MPN impacts which decision variables each constituent mathematical program can influence, either directly or indirectly via solution graph constraints representing optimal decisions for their decedents. Many existing problem formulations can be formulated as MPNs, including Nash Equilibrium problems, multilevel optimization problems, and Equilibrium Programs with Equilibrium Constraints (EPECs), among others. The equilibrium points of an MPN correspond with the equilibrium points or solutions of these other problems. By thinking of a collection of decision problems as an MPN, a common definition of equilibrium can be used regardless of relationship between problems, and the same algorithms can be used to compute solutions. The presented framework facilitates modeling flexibility and analysis of various equilibrium points in problems involving multiple mathematical programs.

3 - Contextual Stochastic Bilevel Optimization

Yao Xie, Georgia Institute of Technology, Atlanta, GA, United States, Yifan Hu, Jie Wang, Andreas Krause, Daniel Kuhn

We introduce contextual stochastic bilevel optimization (CSBO) -- a stochastic bilevel optimization framework with the lower-level problem minimizing an expectation conditioned on some contextual information and the upper-level decision variable. This framework extends classical stochastic bilevel optimization when the lower-level decision maker responds optimally not only to the decision of the upper-level decision maker but also to some side information and when there are multiple or even infinite many followers. It captures important applications such as meta-learning, personalized federated learning, end-to-end learning, and Wasserstein distributionally robust optimization with side information (WDRO-SI). Due to the presence of contextual information, existing single-loop methods for classical stochastic bilevel optimization are unable to converge. To overcome this challenge, we introduce an efficient double-loop gradient method based on the Multilevel Monte-Carlo (MLMC)

technique and establish its sample and computational complexities. When specialized to stochastic nonconvex optimization, our method matches existing lower bounds. For meta-learning, the complexity of our method does not depend on the number of tasks. Numerical experiments further validate our theoretical results.

4 - Hybrid Algorithms for Enhanced Efficiency and Scalability of Network-Based Tri-Level Interdiction Models

Kash Barker, University of Oklahoma, Norman, OK, United States, Nafiseh Ghorbani-Renani, Andres Gonzalez

This study focuses on optimizing resilience strategies for interdependent infrastructure networks (e.g., electric power, water supply, transportation system) against intelligent attacks. We formulate this problem as a tri-level defender-attacker-defender (DAD) model, which is known for its computational complexity. To address this challenge, we propose two novel hybrid decomposition-based algorithms: the hybrid Benders decomposition-based (HBD) algorithm and the hybrid set covering-based (HSC) algorithm. These algorithms efficiently solve the nested subproblems (master problem and subproblem) of the tri-level DAD formulation, incorporating metaheuristic algorithms for improved performance. The proposed algorithms are applied to a tri-level protection-interdiction-restoration model to optimize network resilience. Two case studies are used to evaluate the effectiveness of the solution techniques: (i) an interdependent system of water, gas, and power networks with 125 nodes and 164 links, and (ii) a simulated system of two networks with 252 nodes and 507 links. Our results demonstrate that both hybrid algorithms offer high-quality solutions with significantly improved computational efficiency compared to the existing exact solution method based on the set covering approach.

5 - Optimal Distance Protection in Power Systems

Josh Taylor, New Jersey Institute of Technology, Newark, NJ, United States

Power system protection schemes today rely on large, unbalanced currents to detect faults. In inverter-dominated power systems, a fault current might be roughly balanced and just a few percent larger than normal, making fault detection more difficult. One solution is for the inverter to slightly perturb its output current and/or voltage, i.e., to inject an auxiliary signal, so as to make the system's behavior under faults easier to distinguish from normal. In this paper, we optimize auxiliary signals for fault detection with distance relays.

We formulate a general auxiliary signal design problem with constraints imposed by system operational requirements, and additive and multiplicative noise. We use a relaxation and duality to reformulate the bilevel program as a semidefinite bilinear program. In the special case of additive uncertainty and no constraints, we obtain an analytical lower bound on the magnitude an auxiliary signal must have to guarantee detection. We solve the optimizations in an example based on distance protection, in which the auxiliary signal is negative sequence current.

6 - Learning-to-Optimize for Consolidation and Transshipment in Multi-Store Order Delivery

Xin Wang, HEC Montréal, Montreal, QC, Canada, Okan Arslan, Jean-François Cordeau, Erick Delage

This study investigates multi-store order delivery services where customers can order from multiple stores for home delivery. We first consider separated-order delivery, where orders from different stores are processed and delivered individually. To improve customer convenience and operational efficiency, we introduce consolidated-order delivery, enabling customers to place a single order across stores and receive all items in one combined delivery. While this enhances convenience, it can increase delivery times due to additional routing for visiting multiple stores. To mitigate this shortcoming, we propose a consolidated-order delivery system with transshipment, allowing drivers to transfer orders at transshipment nodes for higher efficiency. We develop a mixed-integer linear program for the multi-store order problem that models different delivery systems, including separated-order delivery and consolidated-order delivery with or without transshipment. Due to computational challenges arising from routing decisions and time variables, we adopt a learning-to-optimize approach that integrates machine learning and optimization. Our experimental study reveals that Graph-based Neural Network consistently performs the best among four learning methods in terms of accuracy, optimality gap, efficiency, and scalability to larger problem instances beyond the training set. We conduct experiments across four U.S. regions in a realistic setting with dynamic customer arrivals. We find that consolidated-order delivery with transshipment, coupled with a short-duration waiting strategy, consistently delivers superior performance, yielding shorter order completion times and reduced driver travel times through effective spatial and temporal consolidation. Waiting longer to batch more customers is advantageous under conditions of frequent customer arrivals and limited driver availability.

SE35

Building B Level 2 B209

Humanitarian and Non-Profit Operations

Invited Session

Public Sector OR

Chair: Reut Nocham, Tel-Aviv University, Tel-Aviv, Israel

1 - Balancing Effectiveness and Equity in the Face of Uncertainty: the Case of Humanitarian Lateral Transshipments

Reut Nocham, Tel-Aviv University, Tel-Aviv, Israel

Humanitarian organizations often struggle to allocate limited resources efficiently while ensuring fair outcomes for disaster-affected communities. This paper addresses the Humanitarian Transshipment Problem (HTP), focusing on the joint optimization of pre-positioned inventory and post-disaster lateral transshipments. The challenge lies in balancing effectiveness (meeting needs) and equity (fair aid distribution), especially under uncertainty. We propose a novel modeling framework for HTP that formally incorporates equity under uncertainty. While earlier studies have explored the effectiveness–equity trade-off, fairness models under uncertainty often lack rigorous justification. To address this, we extend an established social welfare metric to a stochastic setting, linking within-scenario and between-scenario trade-offs. This results in an axiomatic approach that maintains desirable properties across scenarios. Using real and synthetic data, we show our model improves system performance and supports more equitable decision-making under uncertainty. For managers, this framework offers a practical tool to allocate resources more fairly and effectively. By modeling trade-offs at multiple decision levels, it enables more balanced outcomes. Our findings underscore the strategic importance of lateral transshipments, especially in low-budget, high-uncertainty conditions typical of disaster preparedness. While factors like demand correlation may be uncontrollable, operational budgets can be used strategically to boost performance. This work provides concrete guidance for humanitarian organizations aiming to enhance fairness, responsiveness, and efficiency in their operations.

2 - Shelter Location and Evacuation Planning Under Demand Uncertainty: An Adaptive Robust Optimization Approach

Jaehyuk Kim, North Carolina State University, Raleigh, NC, United States, Osman Ozaltin, Maria Mayorga, Merve Bodur

No-notice disasters, such as earthquakes, that occur with little to no warning pose unique challenges to emergency managers and greatly affect vulnerable populations. Emergency managers must develop plans in advance under uncertainty and act quickly when a disaster occurs. We consider the problem of developing an evacuation plan by determining the locations of shelters to open, the number of trailers that establish initial capacities at each shelter, and the assignment of evacuees to routes and shelters. We propose a two-stage adaptive robust mixed-integer optimization model that allows shelter capacities to be adjusted after observing realized shelter demands. In the first stage, the locations of shelters and trailers and the evacuation routes are pre-determined before uncertain shelter demand is known. In the second stage, after the number of evacuees at each impacted location is realized, evacuees are assigned to shelters via planned routes, and shelter capacities are adjusted as needed. The value of the proposed evacuation planning is assessed by comparing it to the case where evacuees select routes according to their preferences. We solve the adaptive robust mixed-integer optimization model using a partition-and-bound algorithm to minimize the maximum total unmet demand. We conduct numerical experiments using the Sioux Falls network and demonstrate the benefits of our model. The results verify the benefit of evacuation route planning and flexible shelter capacity.

3 - An Empirical Investigation of the Trade-off Between Effectiveness and Equity in Resource Allocation Problems

Ohad Eisenhandler, Afeka College of Engineering, Tel Aviv, Israel, Joachim Meyer, Michal Tzur

The trade-off between effectiveness and equity in resource allocation is a key concern in the public and humanitarian sectors. Utilizing an innovative experimental framework, we capture how decision-makers evaluate various allocation approaches from a centralized perspective rather than the individual-focused analyses commonly

found in the literature. As such, our methodology provides first-of-their-kind insights into balancing these often conflicting objectives. The results reveal that integrating both effectiveness and equity leads to solutions that align more closely with respondent preferences than approaches that focus solely on either aspect. Moreover, as opposed to our upfront conjecture, measures with a stronger emphasis on equity were favored, as long as effectiveness was not substantially compromised. This suggests that decision-makers might be more concerned about equity than previously thought. Gini-based measures emerged as the most accurate in capturing these preferences, while the max-min approach was less aligned with them, indicating a potential disconnect between theoretical models in the literature and actual decision-making practices. Additionally, significant variability in individual responses highlighted differing views on the effectiveness-equity trade-off, suggesting the need for adaptable allocation models to accommodate different preference profiles. These findings provide valuable insights for researchers and practitioners seeking quantitative models to navigate the complex trade-offs involved in resource allocation.

4 - A Chance-Constrained Optimization Framework for Building Retrofitting Decisions Under Natural Hazards

Tanmoy Palit, University of Oklahoma, Norman, OK, United States, Saeid Ghasemi, Andrés D. González, Milad Roohi

Natural hazards such as earthquakes present a major threat to the safety, functionality, and economic stability of communities. The increasing frequency and severity of such events have emphasized the importance of proactive planning and strategic investment in mitigation. This study proposes a chance-constrained optimization framework for determining optimal retrofitting strategies for buildings exposed to natural hazard risks. The model accounts for uncertainty by incorporating multiple scenarios and aims to minimize the maximum expected value loss, subject to budgetary constraints and a probabilistic performance requirement. A case study is conducted in Shelby County, TN, where various earthquake scenarios are simulated to evaluate their potential impacts on the buildings of the community. The chance constraint ensures that the total expected loss stays within an acceptable limit in the required percentage of scenarios, allowing for risk-informed decision-making. The model is tested with different chance-constrained thresholds and budgetary performance levels to evaluate its effectiveness. Results demonstrate the model's ability to target high-risk buildings and allocate limited resources efficiently. This work highlights the value of integrating risk analysis and optimization in resilience planning and offers a practical tool for policymakers aiming to enhance community resilience. By helping communities prioritize retrofitting actions based on potential losses and scenario probabilities, the framework supports long-term resilience and disaster preparedness in a data-driven manner. Although applied to earthquake hazards in this study, the model can also be adapted to support resilience planning for diverse types of hazards or communities.

5 - Towards a Generalized Network Optimization Model for Community Resilience Decision Support

Kenneth Harrison, National Institute of Standards and Technology (NIST), Gaithersburg, MD, United States, Tasnim Faiz, William Hughes

Natural disasters, like floods, tornadoes, earthquakes, and wildfires, wreak havoc on a community's interdependent physical, social, and economic systems, or networks. These natural hazards present loadings (e.g., winds, inundation, ground acceleration) that cause network components to fail, in turn triggering outages impairing the social functions of the community (e.g., housing, businesses, and healthcare); importantly, a protracted period of recovery typically follows. NIST has developed several network optimization models to support community resilience decision-making. Recent advances include modeling the voluntary uptake of mitigation actions (e.g., new building retrofits) under incentive programs, addressing the extended implementation period of mitigation decisions, modeling access to healthcare services, and giving decision-makers tools for addressing risk. This talk describes work towards a generalization of the model for application to multiple communities confronting a range of hazards. Various complicating issues are addressed, including network component failure modeling differences across hazards, aligning with the community's other planning objectives, and satisfying multiple stakeholders. Case studies of communities facing various hazards demonstrate this work towards a common decision support framework for community resilience planning.

SE37

Building B Level 2 B211

Efficiency and Robustness in Data-Driven Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Henry Lam, Columbia University, New York, NY, United States

Co-Chair: Tianyu Wang, Columbia University, New York, NY, United States

1 - Learning with Local and Global Adversarial Corruptions

Soroosh Shafiee, Cornell University, Ithaca, NY, United States

We study learning in an adversarial setting, where an epsilon fraction of samples from a distribution P are globally corrupted (arbitrarily modified), and the remaining perturbations have an average magnitude bounded by ρ (local corruptions). With access to n such corrupted samples, we aim to develop a computationally efficient approach that achieves the optimal minimax excess risk. Our approach combines a data-driven cleaning module with a distributionally robust optimization (DRO) framework. We demonstrate that if the data cleaning module is minimax optimal with respect to the Wasserstein loss, solving an optimal transport-based DRO problem ensures a minimax optimal decision. We further provide tractable reformulations for both modules.

2 - Stochastic Optimization with Decisions Truncated by Random Variables

Jie Wang, Georgia Institute of Technology, Atlanta, GA, United States

Stochastic optimization with decisions truncated by random variables has broad applications in revenue management. However, the training objective is generally nonconvex, even when the transformation function is convex. To address this challenge, we develop a convex hull relaxation that provides an exact reformulation under the assumption of positive dependence in the distribution of the random variable. This leads to an infinite-dimensional convex optimization problem over high-dimensional probability measures, where conventional methods suffer from the curse of dimensionality. To ensure computational tractability, we exploit the problem structure and show that it is equivalent to optimizing over marginal distributions. Based on this insight, we propose a finite-particle approximation that breaks the curse of dimensionality and establish its global convergence guarantees. Finally, we demonstrate the superior performance of our algorithm on a booking limit control problem.

3 - Robust Offline Policy Learning Under Covariate Shifts

Bill Tang, University of Southern California, Los Angeles, CA, United States, Cagil Kocyyigit, Phebe Vayanos

We study the problem of distribution shifts in offline policy learning, where the policy training distribution is different from the deployment distribution and may lead to harmful or sub-optimal policy actions at deployment. In real world applications, changes to an allocation system can cause shifts in measured covariates, such as wording changes in survey questions that elicit different responses from individuals experiencing homelessness. As a result, a non-robust allocation policy may incorrectly over or under allocate resources based on the original offline data distribution. Adopting a Wasserstein distributionally robust approach, we learn an allocation policy that is not restricted to any functional form and robust to potential covariate shifts in the population of allocatees.

4 - Optimizer's Information Criterion: A Diagnostics Framework for Data-Driven Optimization

Tianyu Wang, Columbia University, New York, NY, United States, Henry Lam, Iyengar Garud

Different data-driven optimization methods -- varying in model classes and algorithmic design -- can outperform one another across different data instances. This variability underscores a fundamental yet challenging question: how can we reliably evaluate and compare the out-of-sample objective performance of data-driven decisions when standard sample-based evaluations are often biased relative to the truth? We introduce the Optimizer's Information Criterion (OIC), a general, decision-focused approach that extends Akaike's Information Criterion from model adequacy to decision selection. OIC analytically approximates and removes the bias arising from the interaction between model fitting and downstream optimization, delivering reliable performance estimates without repeatedly re-solving optimization problems as in cross-validation and applying more broadly than existing bias-approximation methods. The framework covers a wide range of formulations including empirical and parametric models, their regularized counterparts, and furthermore contextual and risk-averse optimization formulations.

5 - Distributionally Robust Optimization with Incomplete Covariate Information

Ruiwei Jiang, University of Michigan, Ann Arbor, MI, United States, Qingyuan Xu

Modern decision-making systems increasingly rely on contextual features (covariates) to improve optimization under uncertainty. However, in many datasets covariate information is incomplete. This talk proposes a method based on distributionally robust optimization to leverage those incomplete covariates for robust decision. We discuss the computational tractability and statistical guarantees for the proposed approach and demonstrate its performance in comparison with several impute-then-optimize alternatives.

SE38

Building B Level 2 B212

Network Optimization and Applications

Invited Session

OPT: Network Optimization

Chair: Illya Hicks, Rice University, Houston, TX, United States

1 - The Graph Coupling Problem in Quantum Computing

Saber Dinpazhouh, Rice University, Houston, TX, United States, Illya Hicks

We study an operation compilation problem, associated with the hybrid quantum-classical algorithm, QAOA, for the max-cut problem on trapped-ion quantum computers. The goal is to compile the cost Hamiltonian for max-cut QAOA to global coupling operations and individual qubit flips, as opposed to the standard gate compilation. Rajakumar et al have proposed this compilation and showed such a compilation exists for any graph. To mitigate operation error a short sequence of operations is desired. The problem is a low-rank semi-discrete matrix decomposition for the adjacency matrix of a given graph. The lowest possible rank for this decomposition is known as the graph coupling number, $g_{\text{gc}}(G)$, where G is the input graph. They gave a combinatorial construction, named union of stars, with the rank of at most $3n-2$ for any unweighted graph with n vertices. They gave an $\mathcal{O}(m)$ construction for weighted graphs as well. Here we focus on the unweighted graphs. We prove the order-optimality of the union of stars algorithm for unweighted graphs by discovering a family of graphs with a lower bound of $n-1$ on their g_{gc} . Additionally, we decrease the upper bound from $3n-2$ to $2.5n+2$ for any unweighted graph with n vertices. For specific graph families like cliques, perfect matching graphs, paths, and cycles, we find tighter bounds. Additionally, we propose a compact mixed integer program, which is competitive against their method which has exponential size in the input graph. We also discover a connection of the problem with the Hadamard matrices.

2 - Probabilistic Zero-Forcing

Noah Lichtenberg, Rice University, Houston, TX, United States, Illya Hicks

On a given graph $G = (V, E)$ where each vertex can be colored blue or white, a zero forcing set on G describes a starting set of vertices to color blue, where every other vertex is colored white, such that by applying an iterative deterministic coloring rule, all vertices of G will be blue after a number of iterations. Probabilistic zero forcing is a variation of zero forcing where the iterative color changing rule is no longer deterministic but rather probabilistic. One desired quantity in probabilistic zero forcing would be the expected propagation time, which can be seen as the probabilistic analogue to the number of iterations required to color all vertices blue in zero forcing. In this work, computational implementations for finding expected propagation time are explored.

3 - A New Approach for Solving the Mixed Integer PDE Constrained Source Inversion Problem

Damon Spencer, Rice University, Houston, TX, United States, Mattias Heinkenschloss, Illya Hicks

This talk discusses the numerical solution of a source inversion problem with a linear PDE constraint and sources on the right hand side of the PDE. It has applications to optimal borehole selection. Selecting each source is modeled with a binary variable. The PDE solution is expressed as a function of the possible sources. Because the PDE is linear and the sources are in the right hand side, this leads to a 0-1 quadratic knapsack problem. Glover's linearization is applied to equivalently express the 0-1 quadratic knapsack problem as a mixed integer linear program. For the numerical realization, a finite difference discretization and a quadrature rule are applied. The importance of rescaling the problem is demonstrated. Various techniques such as variable fixing and warm starting are applied to explore the special structure of the source inversion problem. The presented approach performs well compared to other mixed integer nonlinear programming solvers from the literature on this problem class.

4 - New Techniques for the Virtual Network Embedding Problem

Avhan Misra, Rice University, Houston, TX, United States, Illya Hicks

Virtual Network Embedding (VNE) is an NP-hard problem with real-world applications in distributed computing, military networks, high frequency trading, and robotic control. It involves optimally mapping virtual request graphs onto an infrastructure graph under various constraints and conditions. Although Mixed Integer Linear Programming (MILP) formulations exist, they often struggle with scalability and optimality in large instances. To overcome these limitations, we explore reinforcement learning (RL) as a scalable alternative for learning effective embedding strategies. Our work addresses multiple VNE variants, including the minimum resource utilization and subdivision containment variants, and demonstrates how RL-based approaches can approximate high-quality embeddings in large-scale network scenarios where traditional methods fall short.

5 - An Improved Algorithm for Solving A Trilevel Network Interdiction Problem to Optimality by Utilizing Suboptimal Solutions

Joshua Ong, Lawrence Livermore National Laboratory, Livermore, CA, United States

We consider the interdiction defense (ID) problem, a three-stage defender-attacker-defender (DAD) problem where the defender and attacker have access to the same set of components. In the first stage a defender mitigates vulnerabilities by protecting a subset of components. In the second stage an attacker removes a subset of assets from the network to cause the maximum disruption. In the third stage the defender stabilizes the system by routing flow. The best response intersection (BRI) algorithm leverages the ID problem structure to obtain optimal solutions an order of magnitude faster than comparable column and constraint generation algorithms. In this talk, we utilize features of off-the-shelf optimization solvers to obtain suboptimal solutions without incurring additional computational overhead to speed up the BRI algorithm. With this additional information, the BRI algorithm converges in fewer iterations while maintaining optimality guarantees. We show the efficacy of the improved algorithm on a power grid interdiction application. Experiments show that the improved algorithm scalability as the combinatorial complexity increases, achieving three to five times faster run times.

SE39

Building B Level 2 B213

Responsible Decision Making with Discrete Optimization

Invited Session

OPT: Integer and Discrete Optimization

Chair: Swati Gupta, Sloan School of Management, MIT, Cambridge, MA, United States

Co-Chair: Jai Moondra, Georgia Institute of Technology, Atlanta, GA, 30332, United States

1 - Modeling Reputation-Based Behavioral Biases in School Choice

Emily Ryu, Cornell University, Ithaca, NY, United States, Jon Kleinberg, Sigal Oren, Eva Tardos

A fundamental component in the theoretical school choice literature is the problem a student faces in deciding where to apply. Recent models have considered a student who is unsure of their strength as an applicant and can apply to at most k schools of varying selectiveness, caring solely about maximizing their admissions outcome (the quality of the school they attend). However, experience suggests that students' decisions are additionally influenced by crucial behavioral biases based on reputational effects: a subjective reputational benefit when admitted to a selective school, and a subjective loss based on disappointment when rejected. Guided by behavioral

economics work on loss aversion relative to expectations, we propose a behavioral model by which a student balances these subjective effects with the quality of the school they eventually attend, changing their choices in interesting and dramatic ways. In particular, where a rational applicant spreads their applications evenly across the spectrum of selectiveness, a biased student applies very sparsely to highly selective schools and instead primarily targets less selective schools, consequently underperforming a rational student even when the rational student is restricted to a sufficiently large upper bound on applications and the biased student can apply to arbitrarily many. Simultaneously, our model is also rich enough to cover a range of different ways that biased students cope with fear of rejection, including occasionally applying to schools that are too selective. As a result of these "mistakes," biased students obtain a quantitatively lower expected admissions outcome.

2 - Fairness of Normalized P-Means for Allocating Goods and Chores

Paritosh Verma, Purdue University, West Lafayette, IN, United States

Allocating items in a fair and economically efficient manner is a central problem in fair division. For additive preferences, the celebrated notion of Nash welfare is known to produce fair and efficient allocations for both divisible and indivisible goods; there is no known analogue for dividing chores. The Nash welfare objective belongs to a large, parameterized family of objectives called the p-mean welfare functions, which includes other notable members, like social welfare and egalitarian welfare. However, among the members of this family, only the Nash welfare produces fair allocations for goods. Incidentally, Nash welfare is also the only member that satisfies the axiom of scale invariance, which is crucially associated with its fairness properties. In the talk, we will define a class of "normalized p-mean" objectives, which imparts the missing key axiom of scale invariance to the p-mean family. We will show that optimizing the normalized p-mean objectives produces fair and efficient allocations when the items are goods or chores, divisible or indivisible. For instance, the normalized p-means gives us an infinite class of objectives that produce (i) proportional and Pareto efficient allocations for divisible goods, (ii) approximately proportional and Pareto efficient allocations for divisible chores, (iii) EF1 and Pareto efficient allocations for indivisible goods for two agents, and (iv) EF1 and Pareto efficient allocations for indivisible chores for two agents.

3 - Metric Distortion of Small-Group Deliberation

Mohak Goyal, Stanford University, Stanford, CA, United States, Ashish Goel, Kamesh Munagala

We consider models for social choice where voters rank a set of choices (or alternatives) by deliberating in small groups of size at most k , and these outcomes are aggregated by a social choice rule to find the winning alternative. We ground these models in the metric distortion framework, where the voters and alternatives are embedded in a latent metric space, with closer alternative being more desirable for a voter. We posit that the outcome of a small-group interaction optimally uses the voters' collective knowledge of the metric, either deterministically or probabilistically. We characterize the distortion of our deliberation models for small k , showing that groups of size $k=3$ suffice to drive the distortion bound below the deterministic metric distortion lower bound of 3, and groups of size 4 suffice to break the randomized lower bound of 2.11. We also show nearly tight asymptotic distortion bounds in the group size, showing that for any constant $\epsilon > 0$, achieving a distortion of $1+\epsilon$ needs group size that only depends on $1/\epsilon$, and not the number of alternatives. We obtain these results via formulating a basic optimization problem in small deviations of the sum of i.i.d. random variables, which we solve to global optimality via non-convex optimization. The resulting bounds may be of independent interest in probability theory.

4 - Stable Many-to-Many Matching in Opportunity-Aware Employment Systems via Data-Driven Integer Optimization

Sepideh Sedghi, Worcester Polytechnic Institute, Worcester, MA, United States, Marcela Vasconcellos, Andrew Trapp

Current employment systems for assisting the over 123 million worldwide displaced persons, as well as other vulnerable groups such as single mothers and older workers, are largely fragmented and manual. This work builds preference-aware, efficient matching methods to secure formal employment for those too often left at the margins. We model the problem as a many-to-many, cardinal preference-based matching problem with ties, incomplete preference lists, and elastic lower and upper quotas. While prior optimization-based approaches focus on one-to-one and many-to-one settings, we propose a novel integer optimization model for the many-to-many stable matching that accounts for both candidate preferences and job requirements, producing actionable recommendations for a job-matching platform. We start by developing data-driven predictive estimates of candidate-job fit probabilities, which we then use as inputs to enable personalized matching of jobs and candidates. We then introduce first-of-its-kind constraint-based representations of stability into our formulation that eliminate justified envy and waste and incorporate flexible matching bounds to improve outcomes for vulnerable individuals. We demonstrate scalability of our methods to solve instances of size beyond even those observed in practice.

5 - Navigating the Fairness Frontier: Portfolios for Optimization and Machine Learning

Jai Moondra, Georgia Institute of Technology, Atlanta, GA, United States, Swati Gupta, Mohit Singh

Many real-world problems involve balancing multiple conflicting objectives, and the appropriate trade-off among these objectives is often unclear. We introduce the notion of portfolios—small sets of solutions that are approximately optimal for any way of balancing the given objectives. We present general techniques for constructing such portfolios and study their application in three settings: (1) facility location, where different groups experience varying travel distances; (2) scheduling, where machines incur heterogeneous loads; and (3) reinforcement learning, where policy impacts differ across stakeholders. In each case, we consider different families of objective aggregation functions—such as Lp norms, ordered norms, and generalized p-means. We propose algorithms to construct portfolios and provide theoretical guarantees relating the trade-offs between approximation quality, portfolio size, and computational efficiency. Experiments on synthetic and real-world datasets show that our approach effectively summarizes the solution space across diverse trade-off parameters, offering interpretable and robust choices for decision-makers.

SE40

Building B Level 2 B214

Demand Driven Service Operations

Invited Session

Service Science

Chair: Muge Yayla Kullu, University of Texas at San Antonio, San Antonio, TX, United States

1 - The Impact of Capacity Limitations on The Value of Segmentation as a Marketing Strategy

Muge Yayla Kullu, University of Texas at San Antonio, San Antonio, TX, United States, Isil Koyuncu Kilci, Dongling Huang

We empirically investigate how differentiation among the products of the same firm in terms of their quality and capacity use impact a firm's product line decisions, and how they relate to its market performance. We challenge the common wisdom in practice that states that "bigger is better" as measured by seat sizes in the airline industry. Following the analytical assertions found in the literature, we posit that this catchphrase may not hold true due to inefficient management of critical resources. We find that an increase in capacity use has a negative impact on the high-end segment offerings and on the overall performance therein. In other words, "size matters," albeit in the opposite direction than what is thought in practice. Our results also supported the significant impact of segmentation on market performance indirectly through the high-end offerings of the firm. Finally, we find evidence for the halo effect in which increased high-end product offerings would have a positive impact on the overall market share of a multi-product firm, but with diminishing returns.

2 - Fuel Tax Increases and Indexing Under Uncertainty: a Risk-Averse Stochastic Optimization Approach

Isil Koyuncu Kilci, The University of Texas at San Antonio, San Antonio, TX, United States, Abhay Lidbe, Olga Bredikhina, Tim Barnett

Transportation funding based on fuel taxes is becoming unviable due to fuel-efficient vehicles, inflation, rising construction costs, and alternative fuels. Indexing fuel taxes offers a potential solution but involves challenges in choosing methods that account for risks and uncertainties in future revenue streams. This study presents a framework to evaluate fuel tax indexing methods with risk considerations and their impact on one-time tax rate increases to achieve targeted revenues over time. The problem is modeled as a two-stage stochastic optimization, where the first stage determines the index choice, and the second stage calculates the optimal initial tax increase. Using Sample Average Approximation, a Monte Carlo method, the study demonstrates convergence to expected values with large samples. A risk-averse version is also analyzed with chance constraints, showing that increased risk aversion leads to higher initial tax increases and shifts in preferred indexing methods. Sensitivity analyses on crude oil prices, alternative fuel adoption rates, and gas price elasticity provide robust insights. Findings highlight how risk acceptance impacts tax policies, assisting policymakers

in selecting effective fuel tax indexing methods to ensure revenue adequacy while managing user burdens. This study offers a practical tool for policymakers to design adaptive and sustainable transportation funding strategies.

3 - Impact of Short-Term Vacation Rentals on Discretionary Spending and Utilization of Auxiliary Services

Haibo Zhu, University of South Carolina, Columbia, SC, United States, Priyank Arora, Necati Tereyagolu, Sriram Venkataraman

In this study, we investigate the impact of short-term vacation rentals (STVRs) on the demand- and supply-side benefits and costs in the area. We aim to empirically measure the impact of STVRs on discretionary spending in the area as well as the utilization of auxiliary government services, including trash pickup and emergency medical services in the area. Further, our goal is to examine the moderating impacts of various characteristics of the listed property on the relationships between STVRs and discretionary spending and utilization of auxiliary government services in the area. For our empirical analysis, we use data from multiple sources, including AirDNA, South Carolina Revenue and Fiscal Affairs Office, and South Carolina Department of Health and Environmental Control.

4 - Pricing and Delivery Lead TIME Policies for Online Retailers

Zumbul Atan, Eindhoven University of Technology, Eindhoven, Netherlands, Bartu Arslan, Albert Schrotenboer, Saeed Poormoaeid

Online retailers face the critical challenge of jointly managing price and delivery lead time decisions to attract demand while maintaining operational efficiency and profitability. Fast deliveries and low prices can boost sales but may strain last mile logistics and reduce margins. This study addresses the retailer's problem of designing effective pricing and delivery strategies in the face of uncertain demand and customer heterogeneity in reservation prices. We examine three policy structures: (1) a static M option policy, where up to M fixed price and lead time combinations are predetermined and dynamically offered based on inventory levels; (2) a dynamic policy, allowing real time adjustment of one or two options per period depending on inventory; and (3) a hybrid policy, combining a fixed base option with a second dynamically adjustable option. We develop exact and heuristic algorithms to optimize these policies and analyze their structural properties. Through numerical experiments, we assess the operational and revenue implications of flexibility in option design. Our results show that offering more options and using dynamic adjustments can significantly enhance performance, particularly in long selling seasons or when inventory holding and goodwill loss costs are high. These insights can guide retailers in balancing simplicity and flexibility to better match customer expectations and supply constraints.

5 - Joint Spectrum Leasing and Admission Control for 5G Spectrum Sharing Networks Using Reinforcement Learning

Demet Batur, University of Nebraska-Lincoln, Lincoln, NE, United States, Romina Raafat, Jennifer Ryan, Mehmet C Vuran

We propose a continuous-time Markov decision process (CTMDP) to model the real-time admission and leasing decisions of a primary user (PU) in a wireless network. The central resource in such networks is radio spectrum, a limited asset whose usage rights are granted to PUs through licenses. PUs are responsible for providing wireless connectivity to their own customer base, but may also choose to lease excess spectrum to secondary users (SUs) to generate additional revenue. However, mismanaged leasing decisions can lead to spectrum shortages, resulting in network slowdown for the PU's customers. When slowdown occurs, the PU must decide whether to admit new customer arrivals—knowing that further congestion will degrade service quality—or to reject incoming requests to maintain service standards. A slowdown penalty is incurred when service quality degrades. Similarly, rejecting new customers leads to a rejection penalty. This problem is modeled as a CTMDP with a large state space and stochastic dynamics. To address the computational complexity, we implement a reinforcement learning (RL) approach using a tailored Q-learning algorithm. This algorithm learns effective admission and leasing policies without requiring full knowledge of the underlying transition probabilities. It incrementally updates value estimates based on observed state transitions and explicitly accounts for both slowdown penalties and rejection costs in the reward structure. The Q-learning algorithm is designed to balance immediate rewards with long-term system performance. Numerical experiments demonstrate that the policies learned via Q-learning closely approximate the optimal strategies derived from value iteration, while providing significant improvements in scalability.

SE41

Building B Level 3 B301

Resource Optimization and Regulation in Healthcare

Invited Session

Health Applications Society

Chair: Shubham Akshat, University of Houston, Houston, TX, United States

1 - Resource-Constrained Healthcare Optimization: Comparing Objectives' Effect on Multi-Patient Treatment Allocation

Yuanxin Zhu, Carnegie Mellon University, Pittsburgh, PA, United States, Alan Scheller-Wolf, Jon Caulkins, Rahul Telang

This study investigates how objective function selection affects treatment allocation in resource-constrained healthcare environments using Markov Decision Process (MDP) modeling. We compare the impacts of welfare maximization, utility improvement, and readmission minimization objectives on resource distribution across diverse patient status. Our MDP framework captures the stochastic nature of patient health trajectories while incorporating limited resource constraints. Results reveal significant variations in allocation patterns and outcomes depending on the chosen objective, with distinct tradeoffs between immediate welfare gains and long-term readmission rates. These findings provide critical insights for healthcare policy design, suggesting that hybrid objective approaches may better balance short-term utility improvements with sustainable reduction in hospital readmissions.

2 - Turning a Blind Eye: Role of Pharmacy and Supplier Distribution Channels in the Opioid Epidemic

Iman Attari, University of Cincinnati, Cincinnati, OH, United States, Jonathan Helm, Jorge Mejia, Robert Wiedmer

The opioid crisis is driven by the oversupply of prescription opioids. While the DEA has blamed large national suppliers for their role in the oversupply of drugs, we show that other distribution channels, such as small suppliers and self-distribution through vertically integrated supply structures also contribute significantly to the crisis.

3 - Evaluating the Effectiveness of Optn Regulations in Organ Transplantation

Michael Olabode, University of Minnesota - Twin Cities, Minneapolis, MN, United States, Saumya Sinha

Organ transplantation is often the only therapy available to patients with end-stage organ failure. In 2024, over 48,000 transplants were performed in the US, while another 67,000 candidates joined the national waitlist. The Organ Procurement & Transplantation Network (OPTN) manages the national transplant system and oversees the operation of transplant programs (i.e., hospitals that perform transplants). The OPTN has historically monitored transplant program performance using 1-year post-transplant survival metrics. Recently, the OPTN introduced a new set of criteria with additional evaluation metrics to improve performance monitoring. Our study investigates the effectiveness and possible limitations of these regulations using a simulated environment. As a first step, we aim to quantify the prevalence of 'false flagging', i.e., the likelihood that a well-performing transplant program is incorrectly flagged and vice-versa. We develop a Monte Carlo simulation framework comprising three phases: (i) preliminary phase, where we estimate patients' pre- and post-transplant mortality; (ii) simulation phase, where we assign patients to virtual transplant programs and generate simulated patient outcomes; and (iii) evaluation phase, where we assess the accuracy of the new metrics. Because true program-level risks are unobservable, we artificially designate certain programs as high-risk and simulate patient outcomes using preliminary regression models. The models are calibrated using national data on adult first-time kidney transplant candidates waitlisted between 2012 and 2019. Our findings aim to substantiate the OPTN's efforts to refine performance metrics, prioritize patient safety, and promote increased utilization of deceased-donor organs.

SE42

Building B Level 3 B302

From Waiting Rooms to Pharma Markets: Algorithmic and Informational Solutions for Equitable Healthcare

Invited Session

Health Applications Society

Chair: Jiatao Ding, Nanyang Technological University, Singapore, Singapore

1 - Improving Operational Design of Genomic Surveillance to Detect Novel Variants of Infectious Diseases

Pankaj Jindal, Indian School of Business, Hyderabad, India, Sarang Deo, Sripad Devalkar

Genomic sequencing-based surveillance (GSS) enables the detection and monitoring of novel pathogen variants at the population level. However, as demonstrated during the COVID-19 pandemic, limited sequencing capacity can delay variant detection and prevent effective public health responses. It is thus crucial to study ways to improve the operational design of GSS, specifically determining the proportion of positive cases selected for sequencing, to effectively utilize the limited sequencing capacity. This allows for quicker detection of novel variants and timely, targeted implementation of public health interventions.

In this paper, we develop an analytical model of GSS in which we abstract disease transmission dynamics and model the genomic sequencing process as a continuous-time Markov chain. This is complemented by a realistic simulation model in which we relax the assumptions made in the analytical model and explicitly model the disease transmission dynamics, using a compartmental epidemiological model, along with the genomic sequencing process. We prove that sequencing a higher proportion of disease-positive samples is not always optimal. While a higher sampling proportion increases the probability of selecting a sample infected with the novel variant for sequencing, it can delay detection due to longer wait times caused by congestion in the genomic sequencing process. Finally, we demonstrate the benefit of utilizing our analytical model as a decision-making heuristic for several real-world scenarios.

Our results have practical implications in public health decision making for effective management of evolving epidemics and can lead to efficient provision and utilization of genomic sequencing capacity and public health interventions.

2 - Optimized Routing of the Mobile Lung Cancer Screening Units

Kiana Amani, School of Industrial and Systems Engineering, University of Oklahoma, Norman, OK, United States, Iakovos Tournazis, M. Soheil Hemmati

Oklahoma ranks 8th in the nation for age-adjusted lung cancer incidence rates (63.3 per 100,000) and 5th for lung cancer mortality (45.3 per 100,000). Low-dose CT (LDCT) scanning remains the most effective screening modality for lung cancer. While the national participation in lung cancer screening programs remains low, Oklahoma suffers from far lower uptake (11% vs. 16% nationally). As a remedy, mobile lung cancer screening programs have been introduced to bring the LDCT to areas with limited access to screening facilities. However, most mobile lung cancer screening programs remain limited to on-demand visits to patients located in remote sites. An alternative approach is to schedule more regular visits, guided by the lung cancer prevalence. In this work, we develop a routing optimization model for a mobile lung cancer screening under uncertainties in screening participation and outcomes. The model is informed by the county-stratified incidence rates from the Oklahoma Central Cancer Registry and geo-spatial characteristics of Oklahoma. We demonstrate the results of our model when applied to counties with limited access to LDCT facilities and quantify the outcomes of the optimized mobile screening program.

3 - The Role of Information Asymmetry Between Brand-Name and Generic Firms in Pharmaceutical Markets: Limit Pricing and The Generic Competition Paradox

Qinquan Cui, UCL School of Management, University College London, London, United Kingdom, Kenan Arifoglu, Dongyuan Zhan

Brand-name firms, leveraging their extensive resources and patent protection, possess more information about consumers compared to generic firms. This *information asymmetry* plays a crucial role in shaping pharmaceutical markets. Moreover, following the entry of their generic alternatives, some brand-name drug prices experience a paradoxical increase (i.e., the *Generic Competition Paradox (GCP)*). However, existing research overlooks the information asymmetry between brand-name and generic firms, failing to provide a comprehensive understanding of pharmaceutical markets. To address this, we develop a game-theoretic model with signaling to analyze interactions between a brand-name firm and a generic firm over two periods (signaling and full-information). We find that the brand-name firm can use *limit pricing* during the signaling period to deter generic entry by charging below its monopoly price, then increase its price and allow entry in the full-information period, leading to the GCP. Consequently, under information asymmetry, limit pricing may arise from the brand-name firm's *anticompetitive* practices, offering an *alternative* explanation for the GCP. While typically resulting in higher drug prices and reduced consumer choice, limit pricing and the GCP can sometimes *benefit* consumers and society. This challenges the prevailing belief that information asymmetry and the GCP are always detrimental, highlighting their nuanced impacts.

4 - Attention is What You Need: The Impact of an Information Support Tool in Addressing Trauma Center Deserts

Zidong Liu, Cambridge Judge Business School, Cambridge, United Kingdom, Feryal Erhun, Houyuan Jiang

Major trauma is a critical medical emergency requiring rapid intervention. In the UK, inclusive trauma networks (TNs) operate under a tiered model, balancing centralized resources with geographical access. However, trauma center deserts—regions with limited access to Major Trauma Centers (MTCs)—worsen health inequalities. To support pre-hospital teams in the triage process, a standardized trauma triage tool was introduced to enhance decision-making, improve awareness of patient conditions, and facilitate rapid medical decisions in emergencies. This study evaluates the impact of this information support tool on system efficiency and patient outcomes, particularly in trauma center deserts.

Analyzing nine years (2013–2021) of trauma submission data from the East of England Trauma Audit and Research Network (TARN), we assess the tool's impact. Our findings indicate that while the triage tool does not significantly affect survival rates, it reduces the average hospital length of stay by 0.5 days. More importantly, it increases good recovery rates by 11% overall, with a notable 17% improvement for patients located more than 45 minutes from an MTC.

These results highlight the critical role of operational decision-support tools in optimizing trauma care and reducing geographical disparities. By standardizing assessments, enhancing healthcare workers' awareness, and improving real-time triage decisions, such tools ensure timely and appropriate treatment, particularly in underserved areas.

Our study provides empirical evidence for policymakers and healthcare providers to enhance trauma system efficiency and address geographical healthcare inequities in trauma center deserts.

SE43

Building B Level 3 B303

Enhancing Healthcare Access: Drone Logistics and Remote Care Solutions

Invited Session

Health Applications Society

Chair: Shakiba Enayati, University of Missouri - Saint Louis, St Louis, MO, United States

1 - Optimizing Spatial Access to Healthcare Services with Emerging Human Mobility Data

Zhiyuan Wei, California Polytechnic State University, San Luis Obispo, San Luis Obispo, CA, United States

Improving spatial access to healthcare facilities is essential for enhancing overall well-being and health outcomes. People living in healthcare deserts are often associated with lower healthcare utilization and increased risks of morbidity and mortality. In the literature, people's access is often conceptualized as proximity, representing their potential access to healthcare services. However, potential access may not reflect people's actual access patterns that reveal how people utilize healthcare services in practice. Thus, to address this gap, this study first extracts human mobility patterns from large-scale mobile phone data to better characterize people's actual access to healthcare services. Then, we develop a data-driven optimization framework that integrates human mobility behaviors into location modeling to improve the spatial

accessibility of healthcare facilities. Specifically, we formulate a choice-based mobility model to capture human mobility patterns, which is then embedded into a facility location model. Finally, we conduct a case study aimed at improving spatial access to healthcare services through optimized facility locations. Overall, this study presents a new perspective on healthcare access by leveraging emerging human mobility data to improve people's actual access patterns to healthcare services.

2 - Predictive Analytics Driven Virtual Appointment Scheduling for a Psychiatry Clinic

Emirhan Uzunlar, University of Nebraska-Lincoln, Lincoln, NE, United States, Ali Dogru, Ibrahim Capar, Ozgur Araz

Telemedicine has improved healthcare accessibility by reducing travel burdens, increasing patient attendance, and enhancing follow-up care, particularly for rural and underserved populations. This study develops a machine learning-based approach to predict whether patients are likely to use telemedicine or in-person appointment options using Medical Expenditure Panel Survey (MEPS) data, focusing on demographics and insurance information. The first phase evaluates various machine learning models, finding that Random Forest achieves the highest AUC (96%) for psychiatry patients. In the second phase, the predicted outcomes inform an optimized scheduling system for a psychiatry clinic, which is evaluated using Sample Average Approximation (SAA) over multiple simulated days. The proposed hybrid appointment system aims to minimize patient waiting time, physician idle time, and overtime, improving overall clinic efficiency. Our computational experiments lead to several critical insights: 1) uneven capacity allocation between service modes can challenge common heuristics, with both smaller variance first (SVF) and larger variance first rules showing advantages in specific a scenario 2) modeling service mode-specific no-show behavior and unpunctuality are crucial for balancing patient experience and resource utilization 3) when a patient group with a longer service duration dominates the schedule, they are scheduled toward the end of the session. Otherwise, scheduling the patient group with a shorter service duration later in the session is more effective.

3 - A Multimodal Optimization Model for Drone-Integrated Organ Transportation Networks

Bahare Askarian, University of Missouri, Columbia, MO, United States, Suchithra Rajendran, Shakiba Enayati

The increasing demand for organ transplants has heightened logistical complexities within healthcare systems, leading to significant losses of viable organs due to transportation inefficiencies. To address these pressing challenges, we introduce a novel multimodal logistics framework combining drones with conventional transport methods to optimize organ delivery and reduce cold ischemia time (CIT), a critical factor contributing to organ wastage. Our research presents an innovative mixed-integer linear programming (MILP) model designed to strategically coordinate vehicle routing, mode selection, timing constraints, and drone energy consumption management specifically for transporting various organ types from donor hospitals to transplant centers. The model captures a detailed network that includes depots, transplant facilities, organ procurement sites, and drone battery-swapping stations. It rigorously accounts for operational constraints such as vehicle capacities, strict delivery timing and schedules, heterogeneous drone capabilities, energy consumptions, and necessary battery-swapping procedures during transit. Validated through a realistic case scenario, our methodology demonstrates tangible improvements in reducing delivery durations, enhancing operational flexibility, and achieving cost efficiencies compared to traditional approaches. By incorporating drone technology into organ logistics, this research significantly advances the operations research literature and provides actionable recommendations for policymakers, healthcare logistics managers, and infrastructure planners. Our findings enable informed decision-making and promote strategic investment in drone-based transport infrastructure, ultimately modernizing and enhancing the effectiveness of organ transportation networks.

4 - Drone-Enhanced Medical Supply Chains in Rural Areas: a Weather-Integrated Optimization Approach

Pardis Bahmani, University of Missouri - Saint Louis, St Louis, MO, United States, Shakiba Enayati, Sina Ansari, Srikanth Gururajan, Ziteng Wang

Rural regions face persistent challenges in healthcare delivery due to limited infrastructure, long travel distances, rugged terrain, and adverse weather. These barriers often delay medical supply transport, reducing the timeliness and quality of care. While drones offer promise for last-mile healthcare logistics, existing models often overlook realistic, context-aware assessments of drone energy consumption under variable environmental conditions. This study develops a multi-modal optimization framework that coordinates drone and ground vehicle operations to enhance rural healthcare logistics. The model jointly determines routing, siting, scheduling, and modal transitions to minimize total system costs while improving flexibility and reliability. It supports transshipment and vehicle-route interoperability, enabling agile logistics in dispersed, resource-limited areas. To account for environmental variability, the service area is partitioned into weather-informed zones using Voronoi tessellation based on nearby meteorological stations. Drone energy consumption is modeled using validated aerodynamic equations that incorporate wind speed and direction. A time-dependent component captures diurnal atmospheric effects, enabling temporally adaptive and weather-aware energy estimation. We demonstrate the model's utility through a case study of Shield Illinois, a statewide COVID-19 testing initiative. Results show that drone integration reduces operational costs, improves efficiency, and enhances the sustainability of healthcare delivery. The framework serves as a robust decision-support tool for reconfiguring rural health supply chains, promoting equitable access, and guiding data-driven investments in emerging healthcare logistics technologies.

5 - Assessment of Primary Healthcare Accessibility in Rural Areas

Valentina Castaneda-Torres, Georgia Institute of Technology, Atlanta, GA, United States, Sofia Perez-Guzman, Valentina Gutierrez-Gutierrez

Providing primary healthcare (PHC) is a global challenge, particularly in rural areas where nearly half of the population lives and lacks adequate access to essential healthcare services. This research addresses the assessment of accessibility to PHC in rural areas, where multiple socioeconomic, geographic, and infrastructure factors limit service accessibility and exacerbate populations' vulnerability. We propose a p-median facility location problem and explicitly model vulnerability by employing deprivation cost functions. We conduct a case study in a rural region in Colombia, and a series of sensitivity analyses of the impact of deprivation and terminal costs on PHC access. The findings show that 22% of the municipalities lacks access to PHC service points, and among those with access, 20% shared PHC service points and ranked highest in deprivation costs. This research also shows that terminal costs highly influence the optimal allocation of PHC service points to populations, conditioning the prioritization of more vulnerable populations when allocating essential resources.

SE44

Building B Level 3 B304

Healthcare Operations and Policies

Invited Session

MSOM: Healthcare

Chair: Mili Mehrotra, University Of Illinois Urbana Champaign, Champaign, IL, United States

Co-Chair: Tharanga Rajapakshe, University of Florida, Gainesville, FL, United States

1 - Understanding Federal and State Authorities' Disagreement over the Allocation of Critical Equipment Inventory During the COVID-19 Pandemic

Diwakar Gupta, University of Texas, Austin, TX, United States, Mili Mehrotra

Early in the COVID-19 pandemic, a disagreement between federal and regional governments erupted in the public arena. It was about the allocation from the federally held stockpile of critical-equipment (CE) inventory, specifically the ventilator inventory. Whereas the regions preferred to know their allocations with certainty, the federal government kept such information ambiguous. We develop a novel game-theoretic model that captures the impact of allocation ambiguity on the interplay between two possible actions by the regional governments, i.e., exerting effort on reducing the spread of infection and purchasing CE inventory on their own. We obtain equilibrium actions (purchase quantities and effort levels) of regional authorities under certainty of allocation and various levels of ambiguity. We show that the regions would always prefer certainty, whereas the federal government may not. We refer to the federal government's choice of optimal level of ambiguity as strategic ambiguity. We also identify special cases in which their preferences will be aligned and illustrate the impact of strategic ambiguity on regions' costs through numerical examples. Our analysis not only helps explain why disagreements arise, it could be used to compare costs of strategic ambiguity with those of alternative approaches that align federal and regional decision makers' preferences.

2 - A Large Language Model Agent Framework for Balancing Physician Workload in Online Health Platforms

Xiguang Liu, University of Florida Warrington College of Business Administration, Gainesville, FL, United States, Xinyuan Wang, Liangfei Qiu

Physicians practicing online face increasing performance challenges due to workload overload from excessive patient requests, leading to degraded treatment outcomes and diminished patient experience. However, studying these issues through real-world data is often constrained by privacy concerns, limited data availability, and the slow,

costly nature of empirical research. To address this, we propose a Large Language Model (LLM) agent-based simulation framework to investigate and optimize physician-patient-platform dynamics in online healthcare environments. This study focuses on three core challenges: (1) factors influencing doctors' patient acceptance decisions, (2) benefits of physician collaboration groups, and (3) platform-level request allocation optimization. The framework deploys LLM agents in two primary roles: (1) Single-agent decision systems simulating doctors' real-time workload assessments; (2) Multi-agent operational systems simulating patient-platform-doctor interactions to optimize consultation request allocation through medical specialization and physician availability. This approach provides implementable solutions to prevent physician overload and enhance care quality. The proposed framework can also be adapted to other scenarios where three key elements coexist: single-expert dominance, collaborative synergy formation, and cross-modal coordination.

3 - Capacity Management for Sudden Demand Surge: a Queuing Approach

Yu Wang, University of North Carolina, Wilmington, NC, United States, David Cho

This research analyzes operational strategies hospitals can adopt to manage capacity when facing sudden overwhelming demand stream in addition to the ordinary demand. We develop an analytical queuing model based on findings from data reported by the State of California Office of Statewide Health Planning and the U.S. Department of Health & Human Services on performance measures such as patient average length of stay for over 300 hospitals in California.

4 - Hasan's presentation

Hasan Md Mahmudul, University of Florida, Gainesville, FL, United States

SE45

Building B Level 3 B305

LLM in Service Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Ruihao Zhu, Cornell University, Ithaca, NY, United States

Co-Chair: Hanzhang Qin, National University of Singapore, Singapore, Singapore

1 - Generative AI in Action: Field Experimental Evidence on Worker Performance in E-Commerce Customer Service Operations

Lauren Lu, Dartmouth College, Hanover, NH, United States, Xiao Ni, Yiwei Wang, TIANJUN FENG, Yitong Wang, Congyi Zhou

In collaboration with Alibaba, this study leverages a large-scale field experiment to quantify the impact of a generative AI (gen AI) assistant on worker performance in an e-commerce after-sales service setting, where human agents provide customer support through digital chat. Agents were randomly assigned to either a control or treatment group, with the latter having access to a gen AI assistant that offers two automated functions as text messages: 1) diagnosis of customer order issues in real time and 2) solution suggestions. Agents exhibited varied gen AI usage behavior, choosing to use, modify, or disregard AI suggested messages. We employ two empirical approaches: 1) an intention-to-treat (ITT) analysis to estimate the average treatment effect of gen AI access, and 2) a local average treatment effect (LATE) analysis to estimate the causal impact of gen AI usage. Results show that the gen AI assistant significantly enhanced both service speed and service quality. Interestingly, gen AI automation did not lead agents to reduce effort; rather, it increased their engagement, evidenced by a higher message volume and agent-to-customer message ratio. Analysis by agents' pretreatment performance reveals that low performers experienced greater improvements in speed and quality, narrowing the performance gap, while high performers saw a decline in service quality, likely because gen AI suggestions fell below their expertise. These findings underscore the potential of gen AI to improve operational efficiency and service quality while highlighting the need for tailored deployment strategies to support workers with varying skill levels.

2 - Can LLM Chatbots Replace Flow-Driven Nlp Chatbots in Customer Service?

Bolin Song, Cornell University, Ithaca, NY, United States, Vishal Gaur, Ruihao Zhu

Flow-driven natural language processing (NLP) chatbots, such as those built on Google Dialogflow, are increasingly used in customer service for automation, reliability, and cost efficiency, but struggle with complex, unstructured customer queries due to rigid intent-matching mechanisms. The rise of large language models (LLMs) promises to transform customer service with fully autonomous chatbots that can flexibly infer user intent, dynamically generate responses using retrieval-augmented generation (RAG), and provide empathetic tone, though this flexibility may introduce response variability and hallucination. We assess these trade-offs in the deployment of LLM technology in customer service by building an LLM chatbot and analyzing its performance drivers compared to an NLP chatbot in collaboration with a large online marketplace. Our offline evaluations with expert judges reveal that the LLM chatbot outperforms the NLP chatbot in factual accuracy, contextual fit, specificity, and tone even when handling queries programmed into the NLP chatbot. It is further capable of synthesizing answers to complex, multiple-intent queries without direct answers in the training set. In our online field experiment conducted over five months, the LLM chatbot significantly increases overall user satisfaction by 53.6%, better sustains longer conversations, and maintains higher user engagement, whereas the NLP chatbot achieves comparable user satisfaction only when users interact with it through clickable buttons rather than free-typing. Critically, the performance of LLM chatbot depends on its design parameters and generation of training data. Our results demonstrate that LLM-based autonomous chatbots significantly improve customer service and emphasize the complementary strengths of both architectures.

3 - Llm for Large-Scale Optimization Model Auto-Formulation: a Lightweight Few-Shot Learning Approach

Ruihao Zhu, Cornell University, Ithaca, NY, United States, Chung Piaw Teo, Hanzhang Qin, xiao JIN, Kuo Liang, Chunwei Yang, Yuhang Lu, Jianming Mao, Shuyi Sun, Congcong Zeng

Large-scale optimization is a key backbone in modern business decision-making. However, the process of building these models is often labor-intensive and time-consuming. We address this by proposing LEAN-LLM-OPT, a LightwEight few-shot leArNing framework for LLM-assisted large-scale OPTimization modeling, which takes a query (a problem description and associated datasets) as input and orchestrates a team of LLM agents to output the optimization formulation. LEAN-LLM-OPT innovatively applies few-shot learning to guide LLMs to build optimization models through reasoning and customized retrieval-augmented generation (RAG) tool usage. Specifically, upon receiving a query, a problem classification agent first determines the type of the problem. Then, a few-shot example generation agent consolidates a set of examples that demonstrate how optimization models are built for problems of the same type step-by-step. Finally, a model classification agent follows these examples to extract relevant information from the input datasets and generate the final output. Extensive simulations validate that LEAN-LLM-OPT attains state-of-the-art modeling accuracy compared to existing methods, especially on large-scale optimization modeling. Additionally, in a Singapore Airlines choice-based revenue management use case, LEAN-LLM-OPT demonstrates its value by achieving less than 1% revenue loss across a variety of scenarios while o4-mini (OpenAI's latest reasoning LLM) usually suffers from at least 55% revenue loss. Along the way, we introduce Large-Scale-OR and Air-NRM, the first set of large-scale optimization model formulation benchmarks. A demo of LEAN-LLM-OPT is available at <https://lean-opt-llm.streamlit.app/>.

SE46

Building B Level 3 B306

Risk Management in Public Policy

Invited Session

Decision Analysis Society

Chair: Katharina Best, The RAND Corporation, Pittsburgh, PA, United States

1 - An Updated Aimpoint Investment Model

Jeremy Eckhause, RAND Corporation, Arlington, VA, United States, Katharina Best

We present an updated version of the Aim Point Investment Model (APIM), a mixed-integer programming model for portfolio-level investment allocation across military technology acquisition programs based on those programs' utility, subject to constraints on the ability to employ the invested funds productively within the required timeframes. We update the model to handle constraints that provide for more balanced investments over each time period.

2 - Quantifying Risk to Future Readiness: a Simulation-Based Framework for Evaluating Global Force Management Decisions

April Cho, Center for Naval Analyses, Arlington, VA, United States

Decisions to commit forces beyond baseline Global Force Management (GFM) requirements —such as extending deployment or shifting areas of operation—can support urgent priorities in the near-term but often carry long-term risks to fleet readiness and sustainability. This research presents a quantitative decision support framework developed for the Department of Defense to assess the impact of such force management decisions on future readiness.

The Operational Schedule Simulation (OSS) Model integrates simulation techniques with machine learning to generate realistic, forward-looking readiness projections that account for real-world constraints such as shipyard capacity and maintenance inefficiencies. Leveraging historical data, current fleet schedules, and force generation policies, OSS models how force management decisions—such as deployment extensions or reassignments— affect key outcomes including ship availability, fleet presence, maintenance delays, and service life over the long-term. This framework enables users to assess operational risk, compare scenarios side-by-side, and explore mitigation strategies ranging from policy adjustments to infrastructure investments.

By providing rigorous, data-driven insights into operational risk and future readiness, the OSS framework enables a better understanding of how today's operational decisions impact future force readiness. It supports strategic planning and high-level decision-making, helping Department of Defense leadership answer critical questions about the long-term implications of short-term force commitments.

3 - Unmanned Fleet Employment Optimization

Arpita Verma, Center for Naval Analyses (CNA), Arlington, VA, United States

The US Navy is increasingly relying on diverse robotic and autonomous systems (RAS) for a wide array of ISR missions. Effectively integrating these systems with varying capabilities into cohesive force packages and determining optimal employment strategies presents a significant challenge for supporting commanders, particularly when demand for surveillance exceeds available assets. This research addresses the critical need to quantify and select the most effective UxS force mix and employment plans for surveillance missions.

To this end, CNA developed two novel optimization models: the UxS-Route Optimization Algorithm for Monitoring (U-ROAM) and the UxS-Search path optimization tool (U-SPOT). U-ROAM generates optimal schedules for ISR missions over discrete locations, maximizing overall effectiveness and equitable coverage across localized sites to be monitored. U-SPOT optimizes patrol paths for multiple UxS types within a defined operating area (OPAREA) to maximize sensor coverage and the probability of detection. The primary value of these tools lies in their ability to provide optimal mission planning and quantify key performance metrics, such as area covered and time monitored. As the complexity of mission planning escalates with the proliferation of UxS, these tools empower operational planners to maximize asset effectiveness. Furthermore, these models serve as powerful analytical tools for exploring diverse operational scenarios and informing acquisition decisions by simulating the impact of different platforms and capabilities. The developed models are geographically and asset-agnostic, offering broad applicability for various missions and platforms across different areas of responsibility.

4 - Systemic risk in the AI era

Anton Shenk, RAND, Washington DC, DC, United States, Jonathan Welburn

This time is different...right? AI means a lot of things to a lot of people. It's the potential for transformative change. It's an existential risk. Somewhere between, it presents a familiar pattern; large models produced by few firms have gained widespread usage and GenAI tools are highly interconnected across systems deeply embedded in AI tech stacks. Shared dependence on vulnerable in critical sectors ranging from finance to healthcare present clear and growing sources of systemic risk. We explore this risk and what policy can do to get ahead of it.

5 - Robust Pathway Analysis of Electricity Investments Under Net-Zero Uncertainties

Erin Baker, Univ of Massachusetts-Amherst, Amherst, MA, United States

We use Robust Portfolio Decision Analysis to evaluate the sustainability of a set of candidate electricity investment pathways under a range of uncertainties and multiple objectives. Results show the importance of considering a range of objectives: considering cost alone or CO₂ alone resulted in just two non-dominated pathways in each case; adding in the consideration of co-pollutants increased the number of non-dominated pathways to six of the nine considered. Wind and solar generation play large roles in non-dominated pathways, while dominated pathways tend to retain more fossil fuel capacity, which can increase costs and emissions. This analysis highlights the importance of considering fuel price uncertainty and, in particular, the possibility of high natural gas prices, which can lead to high co-pollution in otherwise low-polluting pathways. Results illustrate trade-offs between emissions and costs; as well as between CO₂ and co-pollutants, which is largely due to carbon removal use.

SE47

Building B Level 3 B307

Assorted Topics II

Contributed Session

Contributed

Chair: Hiroko Okajima, Nagoya University, Nagoya, N/A, Japan

1 - A Risk-Reduction Approach to Roadway Infrastructure Adaptation Prioritization

Ashkan Zare, George Mason University, Fairfax, VA, United States, Elise Miller-Hooks

Measures to safeguard transportation infrastructure from uncertain hazard events, such as a flooding event, are crucial for public safety, but require significant investment. As it is impractical to try to protect from all possible futures, a risk-based analysis approach for choosing a best protective strategy for adaptation can prove useful. This paper applies a risk-reduction-based methodology to more deeply understand how alternative approaches to mitigating risk via protective infrastructure investments can affect our neighborhoods. A risk measure, based on remaining connectivity to critical facilities, is employed that gives more weight to neighborhoods that have the greatest gap in protection and, therefore, are at greatest risk of being disconnected from critically needed services. The measure is compared against four alternative investment strategies. The analysis is applied to a study area in the Washington, D.C. Metropolitan area under a variety of hazard modelling assumptions. Trade-offs using ten distinct performance metrics are made. The results demonstrate that the choice of risk model in protective investment planning can greatly impact the recommended adaptation plan and which neighborhoods “win” and which “lose.” The method suggested in this paper provides a balanced distribution of risk and may be a more acceptable approach to a community seeking efficient investment prioritization.

2 - Classification of Discrete-Time Queues and Little's Law

Muhammad El-Taha, University of Southern Maine, Portland, ME, United States

We classify discrete-time queues with scheduling rules into classes and use this classification to address an unresolved issue that Little's law does not apply for all these systems. We also discuss the consequences of this classification on systems' characteristics.

3 - Improving the Uefa Champions League Schedule: a Mathematical Model Reducing Breaks from 64 to 4

Nasser Salmasi, SimpleRose, Wilmington, NC, United States

In the new format of the UEFA Champions League, 32 teams compete in a league phase comprising eight rounds, with each team playing one match per round. This study presents a very efficient mathematical optimization model designed to minimize the number of breaks during this phase, while accommodating the relevant constraints.

Applied to the 2024–25 season, the proposed model produces a schedule with only four total breaks—substantially improving upon the official schedule, which contains 62 breaks.

4 - An Application of Self-Organizing Subspace Clustering for Soccer Player Scouting

wenyong xu, Purdue University, West Lafayette, IN, United States, Nga Hin Ben Fong

Subspace clustering or Projective clustering are commonly used in sports data analytics due to their ability to handle high-dimensional player attributes and identify athlete types with different skill combinations. However, most algorithms only focus on linear relationships, and most of the time neglect the non-linear relationships such as interactive effects between attributes (e.g., how speed and acceleration combine multiplicatively) or threshold effects where performance reaches saturation. This study proposes an enhanced self-organizing subspace clustering algorithm that extends traditional self-organizing maps by incorporating deep feature learning layers. The multi-layer architecture first transforms player attributes through deep neural networks to capture complex non-linear interactions, then applies self-organizing map principles with adaptive subspace selection in the learned representation. Given a prototype player, our method identifies the relevant subspace for that player type with budget consideration and recommends similar players by computing distances in the learned non-linear subspace, while visualizing results in an intuitive 2D map with topological features. Our self-organizing subspace clustering algorithm could be effective in making sports-informed decisions by providing coaches and analysts with both accurate player comparisons and visually interpretable insights into what makes players similar, enabling efficient scouting and talent identification from massive player databases containing millions of potential candidates worldwide.

5 - Data-Driven Product Assortment Optimization Under Consumer Return Risk

Dan Dan Guo, University of Science and Technology of China, Hefei, China, People's Republic of

Fast fashion retailers face growing challenges due to high product return rates, short product life cycles, and increasingly volatile consumer preferences. While returns significantly erode revenue and increase operational costs, assortment decisions often continue to rely on heuristics or historical sales data. To address this gap, we propose a data-driven framework for assortment planning that jointly evaluates product attractiveness and return risk. By linking observable product attributes—such as style, color, and fabric—to both sales performance and returns, our approach enables more informed and profitable decision-making. Integrating feature-level predictions into the assortment optimization process, the framework aims to enhance revenue while mitigating the operational impact of returns.

6 - On the Design and Pricing of Add-On Services

Sreya Kolay, University at Albany, SUNY, Albany, NY, United States

Many firms offer add-on services to complement their core product or service. This paper examines how a firm should price such add-on services, and specifically whether to bundle or unbundle an add-on service with the core product. We find that this decision depends on the nature of the add-on service, i.e. whether its impact varies by customer type or by the quality of the core product, as well as whether the cost of providing the add-on service is incurred as a fixed or variable cost. We also extend the analysis to a product line of two products of varying quality and consider whether the add-on service is provided bundled or unbundled with the entire product line or with one of the products.

7 - Designing Fair Systems: Lessons from Gender Disparities in Competitive Settings

Hiroko Okajima, Nagoya University, Nagoya, Japan, Shigeharu Okajima, Tsuyoshi Nihonsugi, Yoshio Kamijo

This study explores how competition design influences gender differences in performance, using data from professional boat racing in Japan. The sport offers a natural setting in which men and women compete under two types of race formats: standard races, where racers make strategic maneuvers during the race, and fixed-lane races, where competitors are assigned a fixed course, minimizing interaction and emphasizing individual skill.

The analysis finds that in mixed-gender standard races, female racers perform worse than in single-gender settings, while male racers improve. However, in fixed-lane races, no significant gender performance gap is observed. This suggests that the performance decline among women is not due to inherent ability differences, but rather the strategic nature of the competition. When races involve high levels of tactical interaction, gender disparities emerge; when strategy is minimized, the gap disappears.

These findings imply that competitive environments themselves can produce performance gaps across genders. Rather than reflecting innate differences, such disparities may stem from how individuals engage with competitive pressures, expectations, or norms embedded in the structure of the contest.

By highlighting the role of competition design, this study contributes to broader discussions about gender equity in performance-based domains. The insights extend beyond sports to settings like education and workplaces, where strategic dynamics may similarly affect outcomes. Carefully designing systems to minimize unnecessary competitive complexity may promote more equitable opportunities for all individuals.

SE48

Building B Level 3 B308

Stochastic Modeling and Decision Making Under Uncertainty

Invited Session

Decision Analysis Society

Chair: Zhengli Wang, The University of Hong Kong, Hong Kong, Hong Kong

Co-Chair: Junfei Huang, The Chinese University of Hong Kong, Shatin, Hong Kong

1 - An Analytical Framework for Assessing the Impact of Policy Changes to the Kidney Allocation System in the US Under Endogenous Patient Choice

Yuwei Zhou, Indiana University, Bloomington, IN, United States, Baris Ata, John Friedewald, Cem Randa, Yue Hu

This paper develops a modeling framework for assessing the impact of policy changes to the deceased-donor kidney allocation system analytically. To do so, we formulate two models as building blocks: a fluid model of the transplant registry and a model of organ availability and geographic sharing. Both models incorporate the dynamic structural model of transplant candidates' accept/reject decisions for organ offers developed in Ata et al. (2020). They also incorporate important features of the kidney allocation system such as patient and donor heterogeneity, the degree of tissue matching, changes in the health status of transplant candidates as they wait, organ quality, cold-ischemia time as well as the geographical sharing structure. Combining the two models, we develop an equilibrium framework and prove that the equilibrium exists. The equilibrium characterization leads to efficient computations for counterfactual studies regarding possible changes to the kidney allocation system that have been considered by policy makers. In particular, we characterize the new system performance, e.g., the wastage, the waiting times for various patient demographics, and the composition of the transplant waitlist after the policy change. Our counterfactual studies rely on transplant data from 2007-2010 and we compare various allocation policies with the policy that was in effect prior to December 2014.

2 - Community Responder Crowdsourcing for Time-Sensitive Medical Emergencies

Weiliang Liu, University of Chicago Booth School of Business, Chicago, IL, United States, Timothy Chan

In parallel with traditional ambulance dispatch, community first responder (CFR) systems utilize mobile applications to locate and alert nearby registered volunteers (i.e., community responders) for rapid intervention in time-sensitive medical emergencies. A crucial decision is the strategic determination of alert recipients and timings. Insufficient alerts risk delayed intervention, while excessive alerts can reduce future responder engagement. This problem is challenging due to the heterogeneity and uncertainty in responders' behavior, which cannot be precisely predicted. Current CFR systems rely on ad-hoc rules for alert decisions.

In this paper, we develop a dynamic programming (DP) approach to model the alert decision problem and a model predictive control (MPC) algorithm that can generate high-quality solutions very quickly. Our MPC alert policy maintains the optimal DP policy's structure while challenging existing alert rules. Unlike existing rules that expand the alert radius over time, both our MPC policy and optimal DP policy strategically narrow the responder pool for consideration as time progresses. Our numerical studies reveal further limitations of existing rules: 1) Many CFR systems issue all alerts at the start, ignoring different travel times, response probabilities, and real-time behavior of responders. However, the optimal number of initial alerts is typically small and varies according to responder characteristics; 2) The common practice of alerting the nearest responders first is not always optimal. An optimal strategy might alert a more distant responder if the higher probability of alert acceptance outweighs the longer travel time. Our simulations using real-world data validate the superiority of our MPC policy.

3 - From Frustration to Engagement: the Role of Ancillary Services in a Queue

Baolong Liu, ShanghaiTech University, Shanghai, China, People's Republic of, Yang Li, Rowan Wang

Waiting in queues often frustrates customers, reduces service provider profits, and diminishes social welfare. Traditional queue management strategies address these issues either by reducing congestion or redistributing waiting costs, but often overlook opportunities to enhance the waiting experience itself. This paper studies an alternative approach: providing ancillary services that create additional customer value during the wait. We explore a fundamental trade-off: while ancillary services improve customer experience, they may also attract more customers and exacerbate congestion. We develop a game-theoretical queueing model featuring two customer segments: Type-P customers, who value only the primary service, and Type-A customers, who derive additional utility from an ancillary service while waiting. Contrary to conventional intuition, our analysis reveals that ancillary services deliver the greatest benefits precisely when primary service capacity is most constrained. Although ancillary services increase overall congestion, they induce customer self-segmentation: Type-P customers, facing higher delays, exit the system and are replaced by Type-A customers who value the enriched waiting experience. This substitution effect enhances system throughput, improves customer surplus, and increases provider profit. We further compare ancillary services to priority queues and show that ancillary services consistently generate higher throughput and, under broad conditions, deliver superior outcomes in terms of profit and social welfare. The results suggest that service providers, especially those operating under capacity constraints, can benefit from implementing ancillary services. By offering value-added experiences during the wait, providers can strategically reshape customer composition, improving throughput and customer welfare without expanding capacity.

4 - The Ergodic Properties and Invariant Measure of a Two-Dimensional Reflected Ornstein-Uhlenbeck Process

Dongzhou Huang, Colorado State University, Fort Collins, CO, United States, Guodong Pang

We consider a two-dimensional reflected Ornstein-Uhlenbeck (ROU) process, that arises as the diffusion approximation for a parallel server network with a randomly split Hawkes arrival process (or a multivariate Hawkes arrival process) in heavy traffic. We study the ergodic properties of the process, including positive recurrence, and rates of convergence in both the total variation distance and the Wasserstein distance. We also provide a numerical scheme based on a Monte-Carlo method to approximate the invariant measure of the process.

SE49

Building B Level 3 B309

Advances in Sustainable Operations

Invited Session

MSOM: Supply Chain

Chair: Tao Lu, University of Connecticut, Storrs, CT, United States

Co-Chair: Sina Khorasani, Vanderbilt University, Nashville, TN, United States

1 - Does Sustainability Certification Promote Sustainability Investment?

Yurong Liang, Wuhan University, Wuhan, China, People's Republic of, Bin Dai, Tao Lu

While widely acknowledged as a market signaling mechanism, sustainability certification also critically functions as an internal governance tool, enabling firms to reduce their exposure to regulatory and reputation penalties. Although these dual effects—risk mitigation and market signaling—are individually recognized, their strategic interplay and collective influence on a firm's fundamental sustainability investment decisions have been less systematically examined. Notably, the incentives generated by these roles may diverge: information from certification could prompt investment scale-back or market exit due to high costs, even as its signaling value encourages increased investment. In this study, we develop a game-theoretic model to investigate how these strategic considerations interact to shape firms' investment behavior, further exploring the firm's trade-offs concerning certification costs, potential reputation gains, and market demand elasticity.

2 - The Impact of Planogram Design and Display Fees on Retail Category Management

Alper Nakkas, The University of Texas at Arlington, Arlington, TX, United States, Yasin Alan, Mumin Kurtulus

We study the impact of planogram design (i.e., placement of products on shelves) and display fees (i.e., fees manufacturers pay retailers for prime shelf space) on a retailer's category management strategy and interactions with national brand manufacturers (NBMs). We consider a game-theoretic model with one retailer and multiple NBMs. Each NBM offers one product. In addition, the retailer has the option to carry a store brand (SB) product. The NBMs determine their wholesale prices and how much they are willing to pay for the prime shelf space. The retailer not only makes assortment and planogram decisions but also sets the quantities for each product in its assortment, leading to market-clearing retail prices. Our analysis leads to three key findings: First, the presence of prime shelf space affects the retail and wholesale prices in the category and may even require the retailer to change its assortment. Second, there are cases in which the retailer uses SB introduction as a strategic lever to intensify the competition for the prime shelf space and thereby increase its revenue from display fees. Third, despite being an additional expense, display fees can increase the NBMs' profits by allowing them to influence the retailer's assortment and planogram decisions in their favor. We discuss the implications of these findings for managers and researchers.

3 - Sustainable Management of a Water Reservoir System Facing Climate Uncertainty

Martin Glanzer, University of Mannheim Business School, Mannheim, Germany, Felipe Caro, Kumar Rajaram

Water scarcity is a growing global issue. It affects even highly developed regions like California, the world's fifth largest economy, where years of severe droughts interspersed with unusually wet seasons have caused societal debates over the management of the water resources stored in the state's extensive reservoir system. This talk presents a strategic optimization model for the long-term, sustainable management of such a system. Introducing a new modeling paradigm based on cycles of stochastic length, defined by the event when all reservoirs are simultaneously full, we avoid the limitations of traditional models and obtain sustainable management policies. To address the unpredictable effect of climate change on future water supply, we adopt a distributionally robust framework where nature chooses adverse inflow distributions. This leads to a stochastic shortest path problem under distributional ambiguity. Using tools from stochastic dynamic programming and aggregation methods, we obtain policy insights and overcome the curse of dimensionality typically associated with systems models. We illustrate our approach with a case study of California's Sacramento River Basin, where we find suboptimality gaps between 3% and 15%, and show that our sustainable policy reduces average cycle shortage costs by 40% compared to the current policy, demonstrating the significant cost-saving potential of our framework.

4 - Optimal Product Return Insurance

Paolo Letizia, University of Tennessee, Knoxville, TN, United States, Gokce Esenduran, Anton Ovchinnikov

In response to the rising cost of returns, some firms have eliminated free return policies, instead offering customers the option to purchase return insurance. This insurance enables customers to return items and receive a refund. It can be offered by either the firm or an external insurer and purchased by the customer either before or after buying the product. We examine these scenarios to determine how to optimally design both the return policy and the insurance structure.

5 - Mergers of Consumer Cooperatives

Sina Khorasani, Vanderbilt University, Nashville, TN, United States, Cigdem Gizem Korpeoglu

We study consumer cooperatives that aim to achieve buyer power and maximize their member-consumers' welfare (rather than profits). We derive conditions under which mergers of consumer cooperatives benefit or harm their member consumers.

SE50

Building B Level 3 B310

Technology-Enabled Service Operations

Invited Session

MSOM: Service Operations

Chair: Wilson Lin, Santa Clara University, Santa Clara, CA, United States

1 - To Digitalize or Not? the Stress Implications of Digital Transformation

Michela Carraro, NEOMA Business School, Reims, France, David Wuttke, Mrunal Mohadikar, Ankit Uphadhyay, Enno Siemsen, Alexandra Wuttke

Chronic stress is a growing concern in operations, especially as workers face pressure to adapt to new tasks under tight deadlines. Digital technologies like augmented reality (AR) promise to reduce this burden, but little is known about their actual impact on worker stress—particularly when such tools are later withdrawn. Building on Conservation of Resources theory, we examine how the availability of digital and traditional resources affects stress in operational tasks. Across three lab studies using both biological (cortisol) and subjective (perceived) stress measures, we find that use of AR reduces stress in both simple and learning-intensive tasks. However, while AR and traditional resources perform similarly in tasks without learning, the removal of AR in learning contexts triggers a sharp increase in physiological stress—whereas removing traditional support does not. Transitioning from digital to traditional resources is equally disruptive, yielding stress levels comparable to a complete loss of support. These findings reveal that while digital tools can ease stress in the short term, they may also undermine resilience if withdrawn later. Our study offers a new lens on the socio-technical tradeoffs of digital transformation in operations and provides actionable insights for technology adoption strategies.

2 - Algorithm-Enabled Decision Support and Worker Learning: A Large-Scale Field Experiment

Youngsoo Kim, Loyola University Chicago, Chicago, IL, United States, Benjamin Knight, Dmitry Mitrofanov, Yuqian Xu

While algorithmic decision support tools are increasingly adopted in workplaces and often deliver immediate performance gains, their long-term impact on workers' learning remains unclear. This paper explores how algorithm-enabled decision support affects learning dynamics through a large-scale randomized field experiment on the Instacart platform. Workers were randomly assigned to a treatment group with algorithmic support or a control group without it. Results show that access to the tool improved immediate productivity (e.g., picking times) and quality (e.g., refund rates). However, reliance on these tools dampened learning from repeated experience, suggesting reduced opportunities for hands-on problem-solving. We also examine how workers perform once algorithmic support is withdrawn. Those previously exposed to the tool initially experienced a productivity drop but then learned more quickly, implying potential for deeper cognitive development if algorithm use is well managed. Heterogeneity analyses reveal that workers with more prior experience or similar background tasks exhibit less learning substitution and gain more from prior exposure to algorithmic tools. Additionally, algorithmic support has stronger learning benefits in urban settings, pointing to a possible rural-urban divide in tool effectiveness. Overall, our findings suggest that while algorithmic tools can enhance short-term performance, their design and implementation must be carefully managed to avoid undermining long-term skill development. Striking the right balance between automation and learning is key to sustainable workforce performance in an increasingly algorithm-driven world.

3 - Human-in-the-Loop Optimization for AI-Generated Content

Fan ZHANG, University of Oxford, skokie, IL, United States

Artificial intelligence (AI) has revolutionized content creation workflows, yet the cognitive principles underlying effective human-AI collaboration remain poorly understood. This study investigates when human feedback most effectively complements AI processing in collaborative content creation. Using a controlled experimental design with 120 content professionals, we systematically varied human intervention timing across three stages: conceptualization, organization, and refinement. Results demonstrate that human intervention at the organization stage produces significantly higher quality content compared to earlier or later interventions. This advantage reflects cognitive complementarity principles where human analytical reasoning optimally enhances AI-gathered information before narrative structuring. The pattern is explained by three mechanisms: intermediate state processing advantage, dual-process cognitive integration, and reciprocal cognitive scaffolding. These findings establish foundational principles for human-AI cognitive collaboration that extend beyond content creation to domains including medical diagnosis, scientific discovery, and education.

SE51

Building B Level 3 B311

Pushing a Paper Over the Edge: How to Turn a Major into a Minor?

Panel Session

Manufacturing and Service Operations Management (MSOM)

Co-Chair: Hallie Cho, Georgetown University, Washington, United States

Co-Chair: Alp Sungu, Wharton, Philadelphia, PA, United States

1 - Moderator Panelist

Alp Sungu, Wharton, Philadelphia, PA, United States

Our panelists (the EIC and 2 DEs at MSOM) will share practical tips for authors navigating a major revision—especially advice that could help them avoid common pitfalls and streamline the revision process.

2 - Moderator Panelist

Hallie Cho, Vanderbilt University, Nashville, TN, United States

3 - Panelist

Georgia Perakis, Massachusetts Institute of Technology, Cambridge, MA, United States

4 - Panelist

Kamalini Ramdas, London Business School, London, United Kingdom

SE52

Building B Level 3 B312

Emerging Challenges and Opportunities in Sustainable Operations

Invited Session

MSOM: Sustainable Operations

Chair: Arzum Akkas, University of Massachusetts Amherst, Amherst, MA, United States

1 - Boosting Sales and Customer Welfare from Premade Foods (Let the Freshest Chicken Fly off the Shelf First)

Dan Iancu, Stanford Graduate School of Business, Stanford, CA, United States, Jaehyuck Park, Erica Plambeck

We examine a grocery retailer's management of a premade food product. The retailer's goal is to maximize a weighted sum of direct profit and customer welfare. Multiple items of the product are produced in batches, and immediately displayed for sale. Considering that each item's quality decreases while it sits on the shelf, the retailer chooses the shelf life, whether to issue items in FIFO or LIFO order, whether to timestamp items, and how to price items. We find that the retailer should use LIFO issuance and timestamp items only if customers are sufficiently heterogeneous. By extending the model, we identify several features -- such as a holding cost, upper bound on the shelf life, age-dependent disposal cost, or customer risk- or loss-aversion -- that can make FIFO optimal. Lastly, we show how a mandate to donate unsold food items (as implemented in several jurisdictions) can motivate a retailer to increase the shelf life, thereby reducing the quality and quantity of donated items.

2 - Achieving Rawlsian Justice in Food Rescue

Gerdus Benade, Boston University, Boston, MA, United States, Aydin Alptekinoglu

We study a problem faced by a national food rescue platform that matches each donation to the first recipient who claims it. Recipients have very different response rates, leading to a few highly responsive recipients claiming the bulk of the donations. We ask whether priority lists, which control when the donation is announced to each recipient, are a remedy for inequitable outcomes. We show that an n-stage priority list, with an individual notification time specific to each eligible recipient, can achieve any desired expected allocation, hence any fairness target, if the donation is non-perishable. Two-stage or binary priority lists, which notify eligible recipients in only two waves, are simpler to implement and administer but offer less fine-grained control over fairness outcomes. For both perishable and non-perishable donations, we give polynomial-time algorithms to find the n-stage and binary priority lists that optimize a class of Rawlsian objective functions -- maximize the minimum value -- representing a notion of fairness that focuses on the worst-off. The key insight that animates optimal priority list design throughout is to give higher priority (and thus more time to claim a donation) to recipients who have received less in the past and to those who were slower in responding to notifications. We prove that this idea can be codified into a simple index by which to rank order eligible recipients. Computational experiments calibrated by real data confirm that even simple, binary priority lists lead to significantly more fair allocations than the existing first-come-first-serve allocation system.

3 - From Deals to Dumps: the Effect of Promotions on Food Waste in Retail

Alexander Hübner, Technical University of Munich, Straubing, Germany, Konstantin Wink, Fabian Schäfer, Sebastian Georg

This study explores the overlooked connection between promotional activities and food waste in grocery retailing. While promotions are widely used to boost sales and store traffic, their unintended consequences on food waste have not been empirically analyzed until now. Drawing on extensive panel data from a major European retail chain—covering 414 products, 173 stores, and over 65,000 promotion events—this research demonstrates that promotions significantly increase food waste. On average, the waste rate doubles from about 2% to 4%, with short-shelf-life products experiencing up to a fivefold rise. The study also identifies cannibalization as a key factor, with nearly half the tested product groups showing increased waste due to sales being shifted from non-promoted to promoted items. Through counterfactual analysis, we estimate that food waste could be reduced by up to 9% if retailers optimize promotion strategies. This includes discontinuing promotions that fail to generate sales uplift, focusing on items with longer shelf lives, and minimizing promotions with strong cannibalization effects. The findings highlight a critical but often ignored trade-off between marketing objectives and sustainability goals. By showing a direct causal link between promotions and food waste, the paper emphasizes the need for a more integrated approach to promotions and operations planning. This work provides practical insights for retailers aiming to balance profitability with environmental responsibility, urging a shift toward more sustainable promotional practices.

4 - Coordination Contracts for Perishable Products with Returns: Models and Challenges

Ulku Gurler, Wharton School, Philadelphia, PA, Turkey, Emre Berk, M. Ali Ulku

Freshness of perishable products, is affected by the floor operations at a store, including produce handling, stocking, and improper ventilation of storage and display areas for ethylene-sensitive products. Poor in-store handling of the product leads to product returns after customer purchase, resulting in additional tangible and intangible costs for both parties. The supplier's brand and overall market share may suffer greatly due to local retailer-specific customer dissatisfaction. Perishability caused by floor operations can be improved with investments in physical infrastructure as well as training and changes in handling practices. Due to the disparities between suppliers and retailers in terms of capital resources, expertise, and market awareness, the centralized and decentralized decisions regarding the optimal level of investment for improvement may differ. Furthermore, store conditions and practices can be difficult and costly for suppliers to monitor and enforce. Therefore, there is a need to better understand the economic coordination mechanisms suitable for such settings. To that end, we consider a dyadic supply chain for a perishable product consisting of a Supplier and a Retailer under several types of coordinating contracts, where we incorporate payments based on the level of freshness-dependent customer returns into the contract design. We provide some analytical results on optimality in various models. Among the issues we discuss in this talk are (i) the conditions under which coordination can exist, (ii) fairness of the profit- / costsharing mechanisms, (iii) optimal levels of investment by different actors, (iv) impact of asymmetry of information, and the challenges in implementation.

5 - Price-Sensitive Supply: the Role of Uncertainty on Acquisition Pricing

Itir Karaesmen, American University, Washington, DC, United States, Ozden Cakici

We study the decisions of a procurement agent who engages with suppliers (markets) to procure a good. The agent decides on the acquisition prices, the unit cost for procuring the good from the markets. The supply is price sensitive and the quantity procured is known after the acquisition price(s) is (are) set. This is typical in remanufacturing where used products are bought from end users. First, we formulate the agent's decision problem as a newsvendor problem in a single market. We analyze the agent's decision on optimal acquisition price first and study the impact of problem parameters and uncertainty (noise) on the optimal acquisition price(s). We show that the agent may or may not increase the acquisition price as market uncertainty increases. However, we show that the acquisition price is monotonic in the noise when the noise changes according to a particular stochastic order that we call "acquisition price-sensitivity ratio". Then, for the two-market case, we provide conditions under which it is optimal for the agent to diversify and procure from both markets. We show that minimum acceptable acquisition prices in each market constrain diversification, preventing buyers from lowering total procurement costs by sourcing smaller quantities at lower prices across multiple markets. In addition, when markets differ in risk, optimal acquisition prices tilt toward the reliable market, though diversification may still occur. The agent should take advantage of diversification to pool risks and to lower unit procurement costs. However, diversification may not be optimal even when the markets are identical.

SE53

Building B Level 3 B313

Design and Analysis of Niche and Emerging Markets

Invited Session

Revenue Management

Chair: Craig Fernandes, University of Toronto, Toronto, ON, Canada

Co-Chair: Raghav Singal, Tuck School of Business at Dartmouth, Hanover, NH, United States

1 - Optimal Capacity and Price Designs Under Ex Ante and Ex Post Theft

Mingliu Chen, The University of Texas at Dallas, Richardson, TX, United States, Feng Tian, Ruiting Zuo

Internal theft poses a significant challenge in retail firms' operations. Owing to a lack of effective monitoring tools, a firm cannot observe every action in daily operations from its employees, providing room for wrongdoings, such as capacity and cash stealing. A common practice is to increase the price of goods to offset the loss in revenue due to the increasing threat of theft. However, we show that such practices are not optimal. We model the internal theft problem in retailing as a principal-agent model, where the principal (firm) contracts an agent (retail manager) for capacity planning and daily sales. The agent is subject to moral hazard and may steal the capacity before demand realization or steal the sales revenue after demand realization. We solve for the optimal capacity, price, and agent's commission decisions to maximize the

principal's utility. We find that capacity and price decisions are not monotone in terms of the severity of moral hazards. In particular, the principal should first decrease and then increase (increase and then decrease) the price (the capacity) when ex post stealing becomes more prevalent. Simply increasing retail prices and shifting the margin to consumers to combat loss in revenue caused by internal theft can amplify the agency problem in some scenarios since it leads to a significant loss in demand and insufficient commission to the agent. Finally, we investigate the sensitivities of price and capacity decisions to demand uncertainties in the presence of moral hazard.

2 - Hiring Under Congestion and Algorithmic Monoculture: Value of Strategic Behavior

Shihan Chen, University of Pennsylvania, Philadelphia, PA, United States, Jackie Baek, Hamsa Bastani

We study the impact of strategic behavior in a setting where firms compete to hire a shared pool of applicants, and firms use a common algorithm to evaluate the applicants. Each applicant is associated with a score that is observed by all firms, provided by the algorithm. Firms simultaneously make interview decisions, where the number of interviews is capacity-constrained. Job offers are given to those who pass the interview, and an applicant who receives multiple offers accept one of them uniformly at random. We fully characterize the set of Nash equilibria under this model. Defining social welfare as the total number of applicants hired, we then compare the social welfare at a Nash equilibrium to a naive baseline where all firms interview the highest-scoring applicants. We show that the Nash equilibrium greatly improves upon social welfare compared to the naive baseline. In addition, the price of anarchy is small, providing further appeal for the equilibrium. We also show that when firms make interview decisions sequentially and each firm takes the best response action, this process converges to an equilibrium when interview capacities are small. However, we show that the task of computing the best response is difficult if firms have to use its own historical samples to estimate it, while this task becomes trivial if firms have information on the degree of competition for each applicant. Therefore, converging to an equilibrium can be greatly facilitated if firms have information on the level of competition for each applicant.

3 - Multi-Armed Bandit Design of Default Donations

Muzhi Ma, University of California, Los Angeles, Los Angeles, CA, United States, Francisco Castro, Scott Rodilitz

The increasing importance of online donations in charitable giving underscores the need for nonprofits to design effective default donation menus that maximize revenue. Despite the widespread use of such menus, considerable debate persists over how to set default amounts to optimize donations. The primary challenge is the vast number of possible menu configurations, which makes identifying the optimal one complex and resource-intensive. To address this, we present a novel approach using a multi-armed bandit (MAB) framework. We model the problem in a bandit setting, where each default menu serves as an "arm" and the rewards correspond to observed donation amounts. We then reformulate the problem as an online longest path problem in a directed acyclic graph (DAG), enabling an approach with minimal assumptions about donor behavior. Furthermore, we transform the problem into a linear bandit setting and implement the Optimism in the Face of Uncertainty for Linear bandits (OFUL) algorithm. In addition, we also develop a misspecified linear bandit algorithm that addresses scenarios where the reward structure is nonlinear and standard linear bandit methods fail, achieving a sublinear regret bound when the maximum misspecification error is correctly estimated. This research offers a robust, model-independent framework that nonprofit organizations can adopt to enhance their online fundraising strategies, minimizing experimental costs while maximizing revenue.

4 - Peer Review Market Design: Effort-Based Matching and Admission Control

Raghav Singal, Tuck School of Business at Dartmouth, Hanover, NH, United States

Peer review is fundamental to academic research, yet its effectiveness relies on reviewers voluntarily exerting costly effort. Current conference mechanisms, such as affinity-based matching, fail to properly incentivize effort, leading to suboptimal outcomes. This study examines how peer review markets can be designed to maximize high-quality acceptances while minimizing low-quality ones. We model a one-sided market where each agent acts as both an author and a reviewer, submitting papers of private quality and deciding whether to exert effort when reviewing. The conference organizer matches papers to reviewers, aiming to optimize market welfare. We first analyze a mechanism that mirrors current practice, denoted Random, which does not link the agents' dual roles. Next, we analyze a Symmetric mechanism, which segments the market based on agents' effort levels. We show that both mechanisms can perform arbitrarily poorly: Random induces no effort, while Symmetric, though an improvement, can be exploited by low-quality authors. To address these issues, we propose Admission Control (AC), which penalizes reviewers who fail to exert effort by rejecting their submissions. Under perfectly observable effort, AC achieves first-best welfare. When effort is noisily observed, AC uses stochastic penalties to ensure incentive compatibility and guarantees at least 1/2-optimal welfare, with simulations demonstrating stronger performance. Robustness analyses confirm the superiority of AC across various extensions, including noisy review outcomes, author biases, and status disclosures through preprints. Our findings highlight the importance of linking the roles of author and reviewer and conclude with actionable strategies for improving peer review markets.

SE54

Building B Level 2 B216

Causal Inference for Social Systems

Invited Session

Revenue Management

Chair: Hannah Li, Columbia, New York, NY, United States

1 - Double Machine Learning for Causal Inference Under Shared-State Interference

Chris Hays, MIT, Cambridge, MA, United States, Manish Raghavan

Researchers and practitioners often wish to measure treatment effects in settings where units interact via markets and recommendation systems. In these settings, units are affected by certain shared states, like prices, algorithmic recommendations or social signals. We formalize this structure, calling it shared-state interference, and argue that our formulation captures many relevant applied settings. Our key modeling assumption is that individuals' potential outcomes are independent conditional on the shared state. We then prove an extension of a double machine learning (DML) theorem providing conditions for achieving efficient inference under shared-state interference. We also instantiate our general theorem in several models of interest where it is possible to efficiently estimate the average direct effect (ADE) or global average treatment effect (GATE).

2 - Discretion in the Loop: Human Expertise in Algorithm-Assisted Decisions

Lydia Liu, Princeton University, Princeton, NJ, United States

As algorithmic decision systems become increasingly integrated into high-stakes domains like education, understanding their real-world impact requires moving beyond predictive accuracy. This talk bridges AI and human expertise through the lens of causal inference and decision-making. We combine causal theory with empirical analyses of a randomized controlled trial of algorithm-assisted college advising to investigate how advisors use discretion to target interventions. We introduce a notion of non-algorithmic expert interventions based on structural causal modeling and develop a quantitative and qualitative framework to test for the value of human judgment in algorithm-supported contexts. Then, we explore how human behavioral responses to predictive models can shape outcomes in experimental evaluations of ADS. We show that design choices may induce cognitive biases and distort treatment effect estimates. Together, these studies highlight the importance of modeling both human behavior and causal mechanisms when deploying and evaluating AI systems for decision making in the real world. Based on joint work with Inioluwa Deborahaji Raji, Sofia Druchyna, Kara Schechtman, and Hannah Li.

3 - Leveraging Causal Inference for Achieving Human&Ndash;Algorithm Complementarity

Jann Spiess, Stanford GSB, Stanford, CA, United States, Bryce McLaughlin

We formulate an algorithmic-design problem that leverages the potential-outcomes framework from causal inference to achieve human–algorithm complementarity without making restrictive ex-ante assumptions about how algorithmic recommendations affect human decisions. We showcase the utility of our framework using an online experiment that simulates a hiring task. We argue that our approach can make sense of the relative performance of different recommendation algorithms and can help design solutions that realize human-AI complementarity. Finally, we leverage our approach to derive minimax optimal recommendation algorithms that can be implemented with machine learning using limited training data.

SE55

Building B Level 3 B314

Advances in Mechanism Design

Invited Session

Revenue Management

Chair: Shixin Wang, Georgia Institute of Technology, Atlanta, GA, N/A, United States

1 - An Isotonic Mechanism for Overlapping Ownership**Jibang Wu, University of Chicago, Chicago, IL, United States**, Haifeng Xu, Yifan Guo, Weijie Su

This paper designs a simple, efficient and truthful mechanism to improve the peer review at large scientific conferences where papers often have multiple authors and authors have multiple papers. The mechanism elicits self-evaluations from authors and leverages this new source of review data to complement traditional peer reviews, notably striking a balance between statistical efficiency and incentive compatibility.

Our approach starts by partitioning conference submissions into disjoint blocks, each sharing a common set of co-authors. We then elicit the ranking of submissions from each author and employ isotonic regression to produce adjusted review scores, aligning with both the reported rankings and raw review scores. We prove that truth-telling by all authors is a payoff dominant Nash equilibrium for any valid partition of the overlapping ownership sets under natural conditions. Moreover, the truthfulness depends on eliciting rankings independently within each block, making block partition optimization crucial for improving statistical efficiency. Despite being computationally intractable in general, we develop a nearly linear-time greedy algorithm that provably finds a performant block partition with appealing robust approximation guarantees. Extensive experiments on both synthetic data and real-world conference review data demonstrate the effectiveness of our mechanism in practical settings.

2 - Optimal Procurement when Bidders Receive Competing Offers**Justin Burkett, Georgia Institute of Technology, Atlanta, GA, United States**, Anton Kleywegt

We study the problem faced by a buyer charged with procuring a service from one of many potential sellers on an online platform before a deadline. The primary challenge faced by the buyer is that the sellers' costs of providing the service evolve over time, as the sellers are assumed to collect stochastic outside offers from competing platforms as time passes. We show that the optimal mechanism for the buyer in this setting interacts with the sellers at just the first and the last time period. In the first period, the sellers pre-purchase a bid for an auction to be run in the last period. In the last period, the sellers each have the opportunity to adjust their bid upwards (to reflect an increase in their costs) before the auction is run. We discuss the potential application of this mechanism to various industries, including marketplaces for truckload transportation and short-term jobs.

3 - Multi-Item Robust Screening with a Maximin-Ratio Objective**Shixin Wang, The Chinese University of Hong Kong, Hong Kong, Hong Kong**

It is generally challenging to characterize the optimal selling mechanism even when the seller knows the buyer's valuation distributions in multi-item screening. A key insight from the robust mechanism design literature is that when only marginal distributions are known, separate selling is robustly optimal under the maximin revenue objective. However, its performance ratio relative to the fully optimal mechanism can be low. In this work, we address robust screening with a maximin ratio objective. We show that when the marginal support of valuations is known, the optimal mechanism remains a separable mechanism, where each item's allocation probability and payment depend only on that item's valuation. Crucially, we refine the allocation and payment rules by leveraging the available joint support information. This separable mechanism can be implemented through randomized pricing for each product individually, ensuring interpretability, ease of implementation, and strong performance. Our framework also applies to settings where the seller knows marginal support information for aggregate product bundle valuations in any partition of the product set. Beyond the independent support ambiguity set, our approach further demonstrates the optimality of randomized bundling mechanisms under a broad class of ambiguity sets.

4 - Robust Pricing for Cloud Computing**Rahul Deb, Boston College, Toronto, ON, Canada**, Dirk Bergemann

We study the robust sequential screening problem of a monopolist seller of multiple cloud computing services facing a buyer who has private information about his demand distribution for these services. At the time of contracting, the buyer knows the distribution of his demand of various services and the seller simply knows the mean of the buyer's total demand. We show that a simple "committed spend mechanism": it provides the seller with the highest profit guarantee against all demand distributions that have the known total mean demand. This mechanism requires the buyer to commit to a minimum total usage and a corresponding base payment; the buyer can choose the individual quantities of each service and is free to consume additional units (over the committed total usage) at a fixed marginal price. This result provides theoretical support for prevalent cloud computing pricing practices while highlighting the robustness of simple pricing schemes in environments with complex uncertainty.

SE56

Building B Level 3 B315

Robust and Responsible Market Design

Invited Session

Auctions and Market Design

Chair: Juba Ziani, Georgia Tech, Atlanta, GA, United States

1 - Centralization Vs Decentralization in Hiring and Admissions**Diptangshu Sen, Georgia Institute of Technology, Atlanta, GA, United States**, Benjamin Fish, Juba Ziani

There is a range of ways to organize hiring and admissions in higher education, as in many domains, ranging from very centralized processes where a single person makes final decisions to very decentralized processes where many people make decisions about who to admit or hire. Decentralized processes can enable individual and collective empowerment, but this may come at the cost of efficiency. With the advent of automated decision making, this question of centralization has a big impact on hiring and admissions, given that automated systems often are easier to implement, or even require, more centralized decision making.

In this paper, we develop a strategic model to explore the impact of the degree of centralization on both the candidates and the hirers, with a focus on university admissions. The model reflects a trade-off between a centralized committee where preferences may not capture individual hirers' preferences, and a decentralized process where individual hirers face extra costs to interview candidates themselves. We characterize when individual hirers prefer the decentralized process over the centralized process as a function of the degree to which the centralized process and hirers' preferences are aligned. We also show that decentralization can have devastating consequences for fairness, leading to major disparities in the likelihood of getting hired across candidates. Our results demonstrate the trade-offs that occur under the question of centralization vs decentralization, and point to how an answer to this question can impose significant harm to people in these systems.

2 - User Altruism in Recommendation Systems**Ekaterina Fedorova, MIT, Cambridge, MA, United States**, Madeline Kitch, Chara Podimata

Users of social media platforms based on recommendation systems (RecSys) (e.g. TikTok, X, YouTube) strategically interact with platform content to influence future recommendations. On some such platforms, users have been documented to form large-scale grassroots movements encouraging others to purposefully interact with algorithmically suppressed content in order to "boost" its recommendation; we term this behavior user altruism. To capture this behavior, we study a game between users and a RecSys, where users provide the RecSys (potentially manipulated) preferences over the contents available to them, and the RecSys—limited by data and

computation constraints---creates a low-rank approximation preference matrix, and ultimately provides each user her (approximately) most-preferred item. We compare the users' social welfare under truthful preference reporting and under a class of strategies capturing user altruism. In our theoretical analysis, we provide sufficient conditions to ensure strict increases in user social welfare under user altruism, and provide an algorithm to find an effective altruistic strategy. Interestingly, we show that for commonly assumed recommender utility functions, effectively altruistic strategies also improve the utility of the RecSys! We show that our results are robust to several model misspecifications, thus strengthening our conclusions. Our theoretical analysis is complemented by empirical results of effective altruistic strategies on the GoodReads dataset, and an online survey on how real-world users behave altruistically in RecSys. Overall, our findings serve as a proof-of-concept of the reasons why traditional RecSys may incentivize users to form collectives and/or follow altruistic strategies when interacting with them.

3 - Incentivizing Desirable Effort Profiles In Strategic Classification: the Role of Causality and Uncertainty

Vasiliki Efthymiou, Massachusetts Institute Of Technology, Boston, MA, United States, Chara Podimata, Juba Ziani

We study strategic classification in binary decision-making settings where agents can modify their features in order to improve their classification outcomes. Importantly, our work considers the causal structure across different features, acknowledging that effort in a given feature may affect other features. The main goal of our work is to understand **when and how much agent effort is invested towards desirable features**, and how this is influenced by the deployed classifier, the causal structure of the agent's features, their ability to modify them, and the information available to the agent about the classifier and the feature causal graph.

In the complete information case, when agents know the classifier and the causal structure of the problem, we derive conditions ensuring that rational agents focus on features favored by the principal. We show that designing classifiers to induce desirable behavior is generally non-convex, though tractable in special cases. We also extend our analysis to settings where agents have incomplete information about the classifier or the causal graph. While optimal effort selection is again a non-convex problem under general uncertainty, we highlight special cases of partial uncertainty where this selection problem becomes tractable. Our results indicate that uncertainty drives agents to favor features with higher expected importance and lower variance, potentially misaligning with principal preferences. Finally, numerical experiments based on a cardiovascular disease risk study illustrate how to incentivize desirable modifications under uncertainty.

4 - Robustness and Transparency in Online Strategic Classification

Kunhe Yang, UC Berkeley, Berkeley, CA, United States

In many real-world settings -- such as college admissions, loan approvals, or hiring decisions -- decision-makers interact with a dynamic population of agents who strategically manipulate their features to influence outcomes. The **online strategic classification** problem models these interactions and addresses the features the challenge of learning through trial and error by observing only the manipulated features, without prior knowledge of the population, their true features, or sometimes even the feasible manipulations. In this talk, we focus on two critical questions: How can decision-makers ensure **robustness** to strategic behavior borrowing techniques from online learning, and how does **transparency** -- revealing or withholding the classifier -- impact performance? We explore these two questions in a generic graph-based manipulation model and show that, contrary to common beliefs, lack of transparency can actually degrade the decision-maker's performance. We offer an algorithmic framework to adapt online learning techniques to be robust to strategic behaviors.

SE57

Building B Level 3 B316

AI and the Future of Digital Engagement

Invited Session

Social Media Analytics

Chair: Sridhar Nerur, The University of Texas at Arlington, Arlington, TX, United States

Co-Chair: Sagar Mahesh Badve, The University of Texas at Arlington, Bedford, TX, United States

1 - The Impact of Emotional Tone on Online Behavior in Social Contexts

Sagar Mahesh Badve, The University of Texas at Arlington, Bedford, TX, United States

This study explores how emotional tone in social media content influences users' intentions to engage in prosocial behavior, particularly in the context of social justice. Focusing on affective responses, the research examines the role of empathy that connects emotional content with user engagement. By using a randomized controlled experiment, the study investigates how variations in emotional tone may activate empathetic responses that encourage supportive actions such as sharing or endorsing content. Grounded in established frameworks, the findings aim to offer insights into the emotional drivers of digital behavior. This research contributes to a growing body of work on the psychological effects of online content and provides guidance for the ethical design of emotionally resonant media. The results are relevant for scholars, content creators, and digital platforms seeking to promote constructive engagement and enhance the social impact of digital communication.

2 - Twitch Raids: Understanding Viewer Engagement in Digital Communities

Roopa Ramesh Desai, The University of Texas at Arlington, Arlington, TX, United States, Jie Zhang, Kay-Yut Chen

In the dynamic realm of online streaming, Twitch stands out as a vibrant platform where content creators engage with audiences in real-time, fostering communities centered around shared interests and interactive experiences. At the heart of Twitch's ecosystem lies the phenomenon of "raiding," wherein streamers direct their viewers to another channel at the end of their broadcast, amplifying visibility and engagement for the raided streamer. This study explores the dynamics of raiding behavior on Twitch, focusing on how raid frequency influences viewership and community engagement. By employing panel data models, we analyze the factors affecting raids and their implications for streamers and the broader Twitch community. Drawing on Social Exchange Theory, which suggests that individuals engage in interactions based on the exchange of rewards and costs, our findings reveal that streamers who receive raids tend to benefit from increased exposure and viewership, which positively influences their engagement and growth within the community. Additionally, we incorporate the Recency effect to model the temporal decay of raid effects, demonstrating that the influence of past raids diminishes over time. Only recent raids significantly affect viewership, with the impact of older raids becoming negligible. These findings provide valuable insights into the strategic decisions streamers make when participating in raids, emphasizing the importance of timely engagement and reciprocal interactions. By integrating these theories, this study provides practical strategies for streamers to navigate raid dynamics and boost community engagement. Ultimately, it helps streamers better understand raids and optimize their efforts to build active, thriving communities on Twitch.

3 - Small Data in the Big Data Paradigm: LLMs and Relevant use Cases

Fay Cobb Payton, Rutgers University, Newark, NJ, United States

Payton, Lui and Yeager (2024) referenced McClellan and colleagues (2017) for their application of time series analysis and forecasting techniques to identify periods of heightened interest in mental health topics on Twitter (now X). The researchers explored Twitter activities focused on depression or suicide, then forecasted the communication trends over the next thirty days using an autoregressive integrated moving average (ARIMA) model. Their analysis found that spikes in Twitter volume following a behavioral health event often last for less than two days, indicating that people circulate mental or behavioral health information associated with heightened periods of interest for a limited time frame.

As we continue to build and use large language models (LLMs), holistic approaches must be embraced in the ecosystem. AI developers will need to train systems and themselves on small data paradigms which are informed by human experiences, choices, knowledge and extend beyond social determinants of health.

This presentation will discuss a use case of the small data paradigm opportunities in evaluating social media data over a three year horizon with over one thousand observations. This use case challenges traditional data interpretation and offers some latitude to address structural, biological pathways, environmental, policy and missing nuisances.

4 - When Language Backfires: the Unintended Consequences of Disruptive Narratives in AI-Assisted Startup Evaluation

Samaneh Samiee, University of Texas at Arlington, Dallas, TX, United States, Sridhar Nerur, Edmund Prater

In the age of digital engagement, startups' communication of their vision is increasingly mediated through AI-analyzed language and platform-based signals. This study examines how the textual characteristics of startup descriptions influence funding outcomes in competitive digital ecosystems. Leveraging NLP and topic modeling techniques on a dataset of 2,176 early-stage companies, we extract AI-derived features such as disruption orientation, lexical diversity, topic entropy, and linguistic distinctiveness. Grounded in signaling theory and optimal distinctiveness theory, our results indicate that at the early stage, startups employing overly disruptive, complex, or unique language are less likely to secure continued investment. In contrast, firms that describe themselves in a manner aligned with competitive linguistic norms are more successful in securing funding rounds. This research contributes to the literature on digital entrepreneurship and language-as-strategy by demonstrating how AI can illuminate the subtle textual cues that shape investor perception and legitimacy in digital funding platforms.

5 - Innovating Against the Odds: Digital Technology as a Catalyst for Female-Led Venture Success

Siddhi Ramanand Nair, University of Texas at Arlington, Arlington, TX, United States

Women-led ventures often face structural barriers stemming from entrenched gender stereotypes, limiting their access to funding, networks, and growth opportunities. While initiatives such as mentorship and policy support have sought to address these challenges, their impact on enabling scalable, innovation-driven ventures remains limited. This study investigates whether the use of digital technologies helps female entrepreneurs overcome such barriers and improve venture outcomes. Drawing on theoretical perspectives that emphasize the inclusive potential of digital transformation, it explores how digital adoption may reshape access to resources and entrepreneurial participation. By focusing on the intersection of gender, digital technology, and venture performance, this research addresses a critical gap in the literature. The study contributes to broader conversations on digital entrepreneurship and gender equity, offering insights for scholars, policymakers, and practitioners interested in fostering more inclusive innovation ecosystems.

SE58

Building B Level 3 B317

Analytics for Supply Chains and Logistics

Contributed Session

Contributed

Chair: Yang Yang, Northeastern University, Shenyang, N/A

1 - A Two-Stage Stochastic Programming Model for Risk Pooling in Integrated Planning and Operation Under Supplier, Manufacturer, and Retailer Uncertainties

Thanaphorn Sornprom, Lally School of Management, Rensselaer Polytechnic Institute, Troy, NY, United States, Chanaka Edirisinghe

Raw material shortages, manufacturing yield variation, and demand fluctuations are three inevitable uncertainties in managing supply chains. To mitigate these challenges, we apply a risk pooling strategy, called postponement, to delay product shipment so as to gain information on some of the uncertainty in a two-echelon manner. An integrated model for two-period operation is formulated as a two-stage stochastic optimization model, where the second-period decisions are adapted to the first-period decisions and realized uncertainty. The main objective of this study is to address questions related to how much value can be gained using partial information on uncertainties by decomposing production and distribution over two periods, and how that value generated for the integrated supply chain is impacted by the degree of uncertainty of the above three sources, and also if demand' correlations at the retailers make the risk pooling effect even more significant.

Using only first and second moment information of the underlying probability distributions, we develop bounds on system performance to address issues related to the potential value generated due to risk pooling effects, the sensitivity of the demand correlations, as well as supplier and yield volatilities.

2 - Distributed Manufacturing System Design

Paula Penagos, University of Missouri - St. Louis, St. Louis, MO, United States, Haitao Li

With the rapid growth of advanced manufacturing and the prevalent use of Internet of Things (IoT), distributed manufacturing systems have been gaining traction in various industries as a new operational paradigm. Unlike the traditional centralized manufacturing system focusing on achieving economies of scale, a distributed system consist of decentralized facilities across regions that both compete for limited resources and collaborate to achieve some system-wide performance metrics.

We introduce a novel modeling framework for Production Planning in Distributed Manufacturing (PPDM), where a parent firm oversees resource allocation and transportation among multiple autonomous manufacturing facilities. Each facility independently optimizes its production to achieve their own (and conflicting) objectives such as profit maximization, market share, or carbon emission reduction. Transportation serves a dual role: resource distribution and customer demand fulfillment in the PPDM.

The model incorporates a *coopetitive* mechanism that blends both cooperation and competition. Facilities operate with asymmetric information, while the parent firm coordinates supply allocation to optimize global performance goals. A game-theoretic procedure is developed to support decentralized decision-making and system-wide optimization.

A numerical example illustrates trade-offs between production efficiency, transportation cost, and customer satisfaction. The results highlight the value of coordination in distributed systems. This framework bridges gaps between production planning, distributed manufacturing, and supply chain coordination by offering a data-driven, decision-support approach. It provides a vehicle for designing agile, resilient, and efficient distributed manufacturing systems with the *coopetitive* feature.

Keywords: distributed manufacturing, production and transportation planning, supply chain coordination, coopetition, game theory, math programming

3 - A GS1 Standard-Based Integrated Information System for EV Battery Manufacturing and Logistics

Seokho Yoon, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea, Republic of, Yuseon Yun, Jaesung Kim, Yujin Song, Nayoung Lee, Namgyo Kim, Sang Jin Kweon

As the demand for electric vehicles (EVs) accelerates, ensuring transparency, traceability, and sustainability throughout the EV battery lifecycle has become a critical challenge. This study addresses this issue by developing an integrated information system based on GS1 standards to support standardized identification and traceability across the EV battery manufacturing and logistics process. The goal is to establish a data-driven framework that enables end-to-end lifecycle tracking from raw material input to final product packaging by assigning and managing GS1 codes at each critical stage of battery production, including electrode fabrication, assembly, activation, and packing. To this end, we design a GS1 code system tailored to key components such as binders, electrolytes, separators, and other materials, in order to enable comprehensive data aggregation and interoperability across manufacturers and suppliers. We then implement and test a prototype platform using sample process data collected from a domestic EV battery production line and validate its effectiveness in improving traceability and operational efficiency.

4 - Behavioral Implications of Costly Information Acquisition in Newsvendor Model: An Experimental Study

Jiayi Lin, Renmin University of China, Beijing, China, People's Republic of, Wanshan Zhu, Jinxing Xie, Xiaobo Zhao

In the era of information abundance, decision-makers navigate diverse channels for acquiring information, each with varying costs. Our study investigates how these costs influence the decision-makers to choose the accuracy of demand signal and to subsequently make order quantity decisions. The accuracy of demand signal is defined as the probability that the demand signal matches the actual demand realization value. In a newsvendor setting, we characterize the optimal accuracy of demand signal and the optimal ordering strategies as a function of the observed demand signal. Using these results as a benchmark, we conduct laboratory experiments to investigate human biases in information acquisition and inventory decisions. The experimental results show that decision-makers perceive information acquisition costs through Shannon information entropy rather than a linear function of information accuracy. Furthermore, we find that subjects are pulled towards the demand signals in inventory decisions, even when the accuracy of demand signal is zero. Based on these findings, we propose a behavioral model that considers Shannon information entropy and mental accounting. The simulations based on this behavioral model match well with the experiment. These results shed light on how human decision-makers acquire and use demand signal information for inventory decisions.

5 - Optimal Inventory Policy Under Information Sharing About Downstream Inventory Position

Byeongkwon Lee, Seoul National University, Seoul, Korea, Republic of, JeongWook Lim, Kun Soo Park

We investigate periodic review inventory management problems for a manufacturer in a decentralized supply chain under information sharing. The manufacturer orders from its upstream supplier and fulfills demand ordered by its downstream retailer. The upstream supplier has infinite capacity, provides a constant order lead time, and charges zero fixed ordering cost to the manufacturer. The downstream retailer faces independent and identically distributed customer demand, orders from the manufacturer according to a given (s, S) policy, and shares both the realized customer demand and its own inventory position with the manufacturer. The manufacturer incurs linear inventory holding and backorder penalty costs applied to overage and underage scenarios, respectively, and aims to minimize the long-run average holding and penalty costs. We propose an exact optimization algorithm to derive the optimal state-dependent base-stock policy. Furthermore, through comparative analysis with the case without information sharing, we provide insights into the value of information.

6 - Learning-Based Branch-and-Price-and-Cut Algorithm for the Scheduling Problem in the Steel Industry

Ying Meng, Northeastern University, Shenyang, China, People's Republic of, Shikui Wang

In this study, we address a cold rolling mill scheduling problem faced by a steel enterprise in China. The problem focuses on developing an optimal schedule by selecting and sequencing coils to meet the demands of the downstream mills. It is an extension of a classical vehicle routing problem with capacity and color constraints, and the objective aims to minimize the total penalty costs between consecutive coils and the setup costs. To solve the problem, we first formulate it into an arc-flow model and a set packing model. Then, we propose two valid inequalities to tighten the lower bound of the set packing model. A label setting algorithm with dominance rule and fathoming rule is designed to solve the pricing sub-problems. Additionally, to accelerate the process of column generation, we propose two properties to reduce the number of sub-problems that need to be solved. Finally, a machine learning-based branching rule is proposed to obtain the optimal solution for the problem. We validated the proposed algorithm's performance on randomly generated instances based on actual production data. The proposed algorithm outperforms the commercial solver CPLEX and can obtain optimal solutions for up to 72 coils with the learning-based branching rule.

7 - A Column Generation Based Algorithm to Steel Coil Assortment Optimization Problem

Yang Yang, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, China, People's Republic of, Lixin Tang

Steel companies usually produce a variety of steel coils to satisfy demand for a large number of sheeted products required by equipment manufacturers. The conflict between the batch production requirements and the various customer demands has made it inevitable to determine the best assortment of coil type and the assignment of these coil types to finished products so as to reduce production and inventory costs while satisfying as many needs as possible. To describe the problem, a mathematical model is formulated in which the expected inventory holding costs is estimated based on risk pooling economies. Then a column generation based algorithm is developed to solve the problem. Finally, the performance of proposed methods is tested based on randomly generated problem instances.

SE59

Building B Level 3 B318

Healthcare Operations II

Contributed Session

Contributed

Chair: Eyyub Kibis, Montclair State University, Montclair, NJ, United States

1 - Addressing Endogeneity in Healthcare Operations Through Copula-Based and Machine Learning Methods

Meera Simon, Indian Institute of Management Kozhikode, Kozhikode, India, Shovan Chowdhury

Measuring cost at the patient level is important for the measurement and improvement of value. Healthcare providers must therefore be able to provide coordinated care with minimal expenditure which requires accurate estimation of cost. Understanding the importance of individual-level factors that determine expenditure, researchers have developed models that consider several customer-centric factors like socioeconomic and psychological factors and used regression models to analyse the effect of customer profiles and other factors on expenditure. It is, however, established that endogeneity is an important problem that arises while using regression, and many impactful academic journals require that the issue of endogeneity be addressed. Though till 2015, the issue of endogeneity was not of immense importance in operations management, the note from the editors of the Journal of Operations Management published in 2018 emphasised the importance of addressing the endogeneity, encouraging models that could incorporate endogeneity. In our study, we identify two endogenous variables that commonly occur in healthcare (namely, Length of Stay (LOS) and total expenditure) and develop a model that can address endogeneity through bivariate copula which is an instrument-free approach that can effectively be used to model variables with suspected endogeneity. To accommodate other variables that are linked directly to the variables under study, we also use machine learning techniques.

2 - Advancing Kidney Transplant Outcomes Through Multi-Agent Exchange Optimization with Hla Compatibility Constraints

Shayan Sharifi, Wayne State University, Detroit, MI, United States, Melike Yildirim

Kidney transplantation remains one of the most effective treatments for end-stage renal disease, yet the global shortage of compatible organs and rising waiting lists highlight the need for more efficient allocation systems. This study presents a novel optimization framework that enhances transplant opportunities through Multi-Agent Kidney Exchange Programs (MKEPs), with a strong emphasis on biological compatibility and fairness. We develop and compare three mathematical models: (1) a baseline KEP model maximizing transplant count under basic blood and PRA compatibility, (2) an enhanced model that integrates a minimum Human Leukocyte Antigen (HLA) score to improve transplant quality, and (3) a multi-agent model that pools donor-recipient pairs across agents (e.g., hospitals, regions), ensuring fairness while maximizing both quantity and quality of transplants. The multi-agent model demonstrates superior outcomes by significantly increasing the number of transplants (up to 125% improvement over baseline) while maintaining stringent HLA thresholds. Realistic simulations and sensitivity analyses reveal key trade-offs between transplant quality and quantity, providing robust insights for policy design in national and international kidney exchange systems. The results highlight the potential of integrated MKEPs to transform organ allocation policy by balancing fairness, efficiency, and long-term clinical outcomes.

3 - Optimal Chemotherapy Scheduling Under Parameter Uncertainty Using Cem Based Policy Learning

Md Samiul Haque Motayed, The University of Texas at Arlington, Arlington, TX, United States, Shuo Wang

Cancer chemotherapy treatment poses significant challenges due to tumor heterogeneity and the adverse effects associated with drug administration. This necessitates intelligent dosing strategies that can balance treatment efficacy with patient safety. In this study, we present a feedback control framework based on the Cross-Entropy

Method (CEM) for generating adaptive chemotherapy policies. Unlike traditional open-loop strategies such as Pontryagin's Maximum Principle (PMP) or global optimization methods like Quantum Particle Swarm Optimization (QPSO), the approach learns closed-loop policies that can respond in real time to variations in tumor dynamics and treatment response. We first demonstrate that a single CEM-trained policy, calibrated using nominal patient parameters, can achieve favorable outcomes under ideal conditions. However, its performance significantly degrades under spatial parameter uncertainties, particularly in tumor growth rate and drug sensitivity. To mitigate this limitation, we propose an ensemble policy learning strategy, where multiple CEM policies are trained on perturbed patient profiles. During treatment, an expert selection mechanism is used to enhance adaptability. Through extensive numerical simulations, we show that the ensemble approach outperforms single-policy counterparts by reducing treatment cost variability and maintaining stable tumor suppression across uncertain patient scenarios. Compared to PMP, our method offers comparable performance while providing the added benefits of feedback structure, computational efficiency, and ease of implementation. The proposed ensemble policy learning framework offers a promising direction for personalized and adaptive chemotherapy scheduling in uncertain biomedical environments.

4 - Mixed-Integer Programming Formulations for Appointment Scheduling in Tandem System

Tengfang Huang, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Taotao He, Guohua Wan

This work addresses the appointment scheduling problem in tandem (multistage) healthcare service systems, where each patient requires a sequence of services across multiple stages. Previous research has established elegant structural properties, such as submodularity and L-convexity, for the single-stage appointment scheduling problem, enabling efficient algorithms for minimizing expected underage and overage costs under general discrete processing time distributions. However, when extended to tandem systems, these properties no longer hold due to the inherent interdependencies between service stages, which significantly complicates both the theoretical analysis and practical solution approaches.

Despite the extensive study of this problem in the literature, the structural complexity of large-scale tandem systems has been largely overlooked, making scalable optimization a persistent challenge. Building on recent advances, we explore the problem's hidden structure and propose a novel mixed-integer programming (MIP) formulation that enables the efficient solution of large-scale tandem system appointment scheduling problems. Our formulation is directly compatible with state-of-the-art optimization solvers, and, by exploiting structural insights, it facilitates practical computation for real-world healthcare applications. Numerical experiments demonstrate the effectiveness of our approach. This work advances both the theory and practice of appointment scheduling in multistage healthcare systems, providing a new framework for optimizing patient flow and resource utilization at scale.

5 - Modeling Emergency Healthcare Strategies and Capacity Planning During Pandemics

Eyyub Kibis, Montclair State University, Montclair, NJ, United States

COVID-19 has infected millions of people globally and caused significant mortality, underscoring the urgent need for effective detection, reporting, and response strategies by national governments. However, when preventative measures are inadequately implemented or delayed, the outbreak can escalate rapidly, placing immense strain on healthcare systems. To address these challenges, this study develops a compartmental dynamic model to simulate the spread of COVID-19 within a community setting. The model captures the transmission dynamics of the disease and evaluates the impact of various emergency policy interventions. By incorporating real-world outbreak data, we validate the model's predictive capability and assess the effectiveness of potential control strategies. Our findings suggest that the timely deployment of makeshift hospitals can significantly reduce patient density in conventional healthcare facilities, enabling a more efficient allocation of limited medical resources. These temporary healthcare facilities offer a flexible and scalable approach to expanding hospital capacity during crisis periods. The proposed model serves as a practical decision-support tool for state and local governments, helping to plan and implement responsive strategies that mitigate the impact of the pandemic and improve public health outcomes.

SE60

Building B Level 3 B319

APS Distinguished Lecture Mohsen Bayati

Invited Session

Applied Probability Society

Chair: Jing Dong, Columbia University, New York, NY, United States

1 - An Introduction to AI Alignment Challenges and Progress

Mohsen Bayati, Stanford University, Stanford, CA, United States

The performance of state-of-the-art AI is typically measured on limited benchmarks, a paradigm that often fails to capture true real-world deployment needs. This gap is rooted in the problem of model underspecification, where models learn shortcuts that lead to unexpected failures. This tutorial uses this concept as a starting point to introduce the field of AI alignment. Following a brief survey of recent techniques, it will focus on the post-hoc control of opaque models, discussing recent progress and the significant research challenges that remain.

SE61

Building B Level 4 B401

Data-Driven Strategies for Streaming Data Monitoring and Diagnosis

Invited Session

Quality, Statistics and Reliability

Chair: Yue Jiang, Purdue University, West Lafayette, IN, United States

Co-Chair: Ana Estrada Gomez, Purdue University, West Lafayette, IN, United States

1 - Prognostics and Diagnostics of Multi-Sensor Systems with Unknown and Unquantified Failure Modes via Bayesian Nonparametric Process Mixtures

Kani Fu, University of Florida, GAINESVILLE, FL, United States, Minhee Kim

To prevent unexpected failures in manufacturing systems, failure mode recognition and prognostics play a crucial role in prognostics and health management (PHM). However, existing methods often rely on strict assumptions, which limit their applicability. First, they typically assume that the number of failure modes is known, which makes it challenging to handle non-stationary and complex systems. Second, they presume that the failure modes of historical units are known, yet in practice, such information is often unavailable, particularly for new systems. Moreover, damaged units are frequently discarded without identifying their primary failure mechanisms. To overcome these limitations, we propose a Bayesian nonparametric framework, which combines the Dirichlet process mixture model (DPMM) and a neural network to simultaneously learn clusters of failure mode and a remaining useful life (RUL) prediction model. Specifically, our framework iteratively updates both the number of failure modes and the parameters of the RUL prediction model, enabling it to adaptively identify dominant failure mechanisms and improve prediction accuracy. Experiments on aircraft gas turbine engine datasets and simulation datasets are presented to demonstrate the framework's superiority in capturing failure modes and RUL prediction accuracy.

2 - Interactive Resource Allocation and Anomaly Detection via Multi-Agent Reinforcement Learning

Xin Xia, University of Wisconsin-Madison, Madison, WI, United States, Jie Guo, Yujie Ma, Chen Zhang, Kaibo Liu

The proliferation of sensor technologies has enabled real-time monitoring across various domains. However, resource constraints—such as limited bandwidth, energy, and processing capabilities—pose significant challenges for continuous data collection and timely anomaly detection. These constraints necessitate intelligent adaptive sampling strategies that select which data streams to monitor at each time step, balancing the trade-off between detection accuracy and operational costs. This study addresses the challenge of decentralized anomaly detection in systems characterized by a hierarchical System–Region–Stream (SRS) structure. We reformulate the problem as a leader-follower decision-making process, wherein a central controller allocates monitoring resources across regions, and each region autonomously selects data streams for local anomaly detection. To optimize this hierarchical decision-making, we develop a scalable multi-agent deep reinforcement learning (DRL) framework that learns long-term policies, effectively mitigating the curse of dimensionality inherent in large-scale systems. Unlike existing methods that often overlook operational costs, our approach explicitly incorporates these costs into the decision-making process, ensuring a more practical and cost-effective monitoring strategy. Extensive simulation studies and a real-world case study validate the efficacy and applicability of the proposed framework, demonstrating its potential to enhance anomaly detection while optimizing resource utilization.

3 - Cost-Efficient Streaming Estimation via Multihypothesis Sequential Transfer

Qunzhi Xu, New York University, New York, NY, United States

We introduce an adaptive sequential transfer learning framework for real-time parameter estimation in streaming environments with multiple candidate source models. Our approach consists of two stages: first, we perform multihypothesis sequential testing to identify the source distribution closest to the target stream; second, we estimate the target parameter using a penalized estimator anchored at the selected source. This procedure balances bias from source mismatch with variance due to limited target data, enabling accurate and efficient estimation under cost constraints. We provide theoretical guarantees under Gaussian designs and demonstrate the method's performance through simulations, highlighting its utility for dynamic monitoring and diagnosis in nonstationary settings.

4 - Rampseg: a Robust Calibration-Free Image Segmentation Method for Reliable Melt Pool Analysis in Laser Powder Bed Fusion

Nazmul Hasan, The University of Arizona, Tucson, AZ, United States, Eung-Joo Lee, Andrew Wessman, Mohammed Shafae

Coaxial monitoring of the melt pool, the molten region formed during laser exposure in laser powder bed fusion (LPBF) additive manufacturing, is critical for ensuring part quality, as its morphology reflects process stability. However, melt pool images are inherently stochastic due to variations in processing conditions, imaging setups, and transient phenomena such as spatter. The impact of this variability on segmentation accuracy has not been thoroughly explored in the literature, resulting in a critical gap in the development of robust methods. As a result, existing segmentation techniques often lack the generalizability and reliability needed to perform consistently across diverse conditions. In response, we propose *RAMPSEg* (Robust Adaptive Melt Pool Segmentation) algorithm, an intensity-agnostic and calibration-free edge-detection-based method that introduces three key innovations: (i) data-driven optimization of edge detection parameters, (ii) a quantitative segmentation review using an edge-to-area ratio to guide refinement, and (iii) an adaptive smoothing feedback loop. Unlike prior methods, *RAMPSEg* avoids arbitrarily selected parameters, over- or under-segmentation, and fixed heuristics. Instead, it achieves self-regulating segmentation that dynamically balances noise reduction and boundary preservation across diverse imaging conditions, enabling more generalizable melt pool analysis. We comprehensively evaluated its effectiveness using three distinct datasets encompassing diverse process conditions, materials, machines, and imaging systems. This unprecedented evaluation demonstrated consistently high segmentation accuracy (~90%) against expert-annotated ground truth, significantly outperforming conventional thresholding, Otsu's method, and state-of-the-art zero-shot vision transformer models (SAM and CLIPSeg). *RAMPSEg* enables reliable melt pool segmentation under dynamic conditions, supporting accurate anomaly detection, process optimization, and quality assurance.

SE62

Building B Level 4 B402

Statistical Models with Manifold-valued Random Objects

Invited Session

Quality, Statistics and Reliability

Chair: Xiaoyu Chen, University at Buffalo, Buffalo, NY, 14260, United States

Co-Chair: Yingyan Zeng, University of Cincinnati, Cincinnati, OH, United States

1 - Differentially Private Geodesic and Linear Regression

Cheng Wang, University of Massachusetts Amherst, Amherst, MA, United States, Carlos Soto

In statistical applications it has become increasingly common to encounter data structures that live on non-linear spaces such as manifolds. Classical linear regression, one of the most fundamental methodologies of statistical learning, captures the relationship between an independent variable and a response variable which both are assumed to live in Euclidean space. Thus, geodesic regression emerged as an extension where the response variable lives on a Riemannian manifold. The parameters of geodesic regression, as with linear regression, capture the relationship of sensitive data and hence one should consider the privacy protection practices of said parameters. We consider releasing Differentially Private (DP) parameters of geodesic regression via the K-Norm Gradient (KNG) mechanism for Riemannian manifolds. We derive theoretical bounds for the sensitivity of the parameters showing they are tied to their respective Jacobi fields and hence the curvature of the space. This corroborates recent findings of differential privacy for the Fréchet mean. We demonstrate the efficacy of our methodology on the sphere, $\mathbb{S}^2$ and, since it is general to Riemannian manifolds, the manifold of Euclidean space which simplifies geodesic regression to a case of linear regression. Our methodology is general to any Riemannian manifold and thus it is suitable for data in domains such as medical imaging and computer vision.

2 - Dynamic Shape State Space Modeling with Regime Switches

Daniel Qiu, Department of Industrial and Systems Engineering, University of Washington, Seattle, WA, United States

We present a novel dynamic state model for shape evolutions with regime changes, where the shape of an object evolves based on control inputs, and the evolution dynamics switch intermittently at unknown instances. Our primary motivating application is modeling shape evolution in additive manufacturing. This research problem is challenging due to the non-Euclidean nature of the shape space, as well as the non-linearity and regime switches in modeling shape sequences. Explicitly modeling this incurs a complicated model and requires estimating a large number of parameters. Therefore, we seek a local approximate estimation for the regime-switching shape dynamics. We establish algorithms and derive open form expressions for parameter estimates. The new approach is tested using benchmark datasets generated from control experiments in additive manufacturing.

3 - Statistical Quality Control with Distributions-Valued Samples

Yingyan Zeng, University of Cincinnati, Cincinnati, OH, United States, Yujing Huang, Xiaoyu Chen

Shewhart control charts were designed to monitor the change in moments (e.g., sample mean, sample standard deviation, etc.) of the observations in a sample. While the natural method to represent the randomness of such observations is based on the probability distribution they follows, limited by the Euclidean spaces where conventional statistics are defined, existing control charts cannot directly monitor probability measures that belong to a metric space referred as 2-Wasserstein space. We argue that moments are insufficient to represent the shape information of the distributions, hence proposing to directly monitoring the samples as random distributions in 2-Wasserstein space. We focus on the optimal transport map T_i which transforms distribution-valued sample μ_i to an in-control Fréchet barycenter $\bar{\mu}$. We prove that T_i preserves an isometric expansion onto a finite set of orthonormal basis functions, which allows us to monitor the change of distribution by the change of basis coefficients over time. Leveraging this theoretical results, we construct the Phase II statistics Hotelling's T^2 and Squared Prediction Error (SPE) via Functional Principle Component Analysis. This framework offers a fundamentally new paradigm for process monitoring while retaining fully distribution-free control limits.

4 - Mixed Distribution-in-Distribution-Out Regression

Xiaoyu Chen, University at Buffalo, Buffalo, NY, United States

A few recent studies reported the linear regression models with distribution-valued samples, which have elevated the regression analysis from Euclidean space to Wasserstein space equipped with 2-Wasserstein metric. However, data collected from real systems are usually a mixture of uncertain measurements represented as

probability distributions and certain measurements represented as scalar values. For example, a patient in a standard electronic health record (EHR) system can produce a distribution of blood pressures measurements, and a scalar variable of his/her gender. Such mixed data pose significant challenges to the current regression models. In this research, we proposed a mixed distribution-in-distribution-out regression model with both distributional and Euclidean covariates.

SE63

Building B Level 4 B403

Predictive Maintenance and Anomaly Detection

Contributed Session

Contributed

Chair: Neda Zarayeneh, Hitachi America

1 - A Comparison of Deep Reinforcement Learning and Neural Network Based Predictive Maintenance

Jasper Ng, Proof School, San Francisco, CA, United States, Maharshi Dhada

Effective maintenance planning is crucial for industrial operations, minimizing costly failures and resource consumption. This paper introduces and evaluates a novel Deep Reinforcement Learning approach for maintenance scheduling. Instead of traditional statistical or Machine Learning-based Remaining Useful Life estimates to schedule maintenance, this method feeds raw time-series sensor data directly to a Proximal Policy Optimization agent, allowing it to learn optimal maintenance policies by leveraging the complete sensor information. The researchers tested this approach in a simulated multi-machine shop-floor environment using the Commercial Modular Aero-Propulsion System Simulation Jet Engine Simulated Data. The Reinforcement Learning agent was compared against a baseline strategy employing a Recurrent Neural Network to predict time until failure and trigger maintenance accordingly. The results demonstrate the significance of the Deep Reinforcement Learning approach: the agent achieved a substantially lower and more consistent operational cost (48.06 ± 2.66) compared to the Recurrent Neural Network baseline (122.085 ± 34.856) in a simulated shop-floor scenario containing unseen test data. This highlights that continued improvement of Deep Reinforcement Learning methods (e.g. bypassing intermediate Remaining Useful Life predictions) can lead to much more effective and reliable maintenance decisions.

2 - Partial Trajectory Analysis for Diagnosing Sequential Manufacturing Processes

Tsuyoshi Ide, IBM T.J. Watson Research Center, Yorktown Heights, NY, United States, Kohei Miyaguchi

When a product is manufactured through a sequence of individual process steps, such as in semiconductor manufacturing, identifying steps responsible for detected quality issues is challenging due to process heterogeneity, combinatorial complexity, and unpredictable fluctuations in individual steps, among other factors. As a machine learning problem, this cross-process attribution can be formalized as a sequence attribution task, where the goal is to compute the attribution score for each step in the process sequence, given an observed pair of process trajectory X and quality metric y .

While input attribution has been studied extensively in explainable artificial intelligence, few studies address inputs with a sequential structure. To address this limitation, we propose a new framework called Partial Trajectory Analysis (PTA). The key idea of PTA is to quantify the influence of each process on y by comparing prediction outcomes for partial process trajectories, including and excluding the target process step. For a fair comparison, we solve an optimization problem such that the downstream trajectory follows the optimal path.

We validate the PTA framework using real-world semiconductor fabrication data from a state-of-the-art front-end-of-line process.

3 - Censored Contextual Bandits for Resource-Constrained Maintenance Scheduling

Yipu Yao, Durham University Business School, Durham, United Kingdom, Li Ding, Yanlu Zhao

We consider a preventive maintenance scheduling problem for multiple machines under budget and manpower constraints, where failure data are subject to censoring. Each machine has an unknown failure distribution; maintenance incurs fixed costs, while unplanned failures lead to substantial penalties. The manufacturer's objective is to find an optimal maintenance schedule to minimize the failure losses under limited resources. Since any failures that occur outside the maintenance windows are unobservable, decision-maker must learn from censored feedback. We formulate the problem as a contextual bandit, where each arm corresponds to a maintenance interval and planning tier pair. We then propose a robust learning algorithm that adapts to censored observations and provides finite-sample guarantees and near-optimality under mild conditions. To accelerate convergence, we introduce a global observable boundary condition shared across all machines, which effectively reduces the arm space. Within this framework, we design a decentralized UCB-based scheduling policy that tailors maintenance actions to individual machine characteristics. Numerical experiments demonstrate the effectiveness and robustness of our approach, outperforming benchmark policies in both expected cost and worst-case scenarios.

4 - A Dual-Ema Mean-Teacher for Semi-Supervised Virtual Metrology Under Process Nonstationarity and Variability

Kyeonghui Kim, Yonsei University, Seoul, Korea, Republic of, Chang Ouk Kim

Virtual metrology (VM) often faces limited labels and frequent process nonstationarity and operational variability due to equipment condition changes and preventive maintenance (PM) as well as drift/shift. We propose a dual-exponential moving average (EMA) mean-teacher, which runs two teachers with different EMA momenta (fast/slow) in parallel. Their predictions on weakly augmented inputs are ensembled to form pseudo-targets: the fast teacher reduces tracking lag after changes, while the slow teacher provides a stable reference that damps noise, yielding targets that are simultaneously less delayed and less jittery. The student is trained with supervised mean squared error (MSE) on labeled batches and output-level consistency between strongly augmented student outputs and the ensembled teacher targets on unlabeled batches. Experiment results demonstrate robustness in terms of performance stability and post-change recovery even under the label scarcity and drift/variability commonly encountered in industrial virtual metrology.

5 - Damdg: Degradation-Aware Multi-Domain Generalization for Unseen Hdd Failure Prediction

Jisu Woo, Yonsei University, Seoul, Korea, Republic of, Chang Ouk Kim

Hard Disk Drives (HDDs) are essential components in large-scale storage systems and are equipped with Self-Monitoring, Analysis, and Reporting Technology (SMART), which continuously collects various indicators reflecting disk health. Although machine learning-based approaches for HDD failure prediction using SMART data have been extensively studied, two key challenges remain in real-world applications. First, traditional machine learning approaches typically assume that the SMART data distributions of healthy and failed disks are similar, and that sufficient labeled data are available for training. However, in practice, these distributions vary significantly across HDD models—even among those from the same manufacturer—and previously unseen models are continuously introduced. Second, in failed HDDs, failure symptoms tend to occur briefly and sporadically amid long periods of normal operation, making them both rare and difficult to distinguish from normal patterns. This sparsity and interleaving nature of failure signals hinder the model's ability to learn them effectively. To overcome these challenges, we propose a Multi-Instance Learning (MIL)-based framework that segments each HDD time series into multiple instances and focuses on the most informative segments. Furthermore, we integrate a Multi-Domain Maximum Mean Discrepancy (MMD) loss into an Adversarial Autoencoder (AAE) architecture to minimize distributional shifts across HDD models and to learn domain-invariant representations. Experimental results demonstrate that our method achieves superior failure detection accuracy and improved generalization to unseen HDD models compared to traditional machine learning approaches.

6 - Forecasting Manufacturing Inefficiencies with External Risk Indicators and Data-Driven Modeling

Muhammad Omar Al-Zadid, Washington State University, Pullman, WA, United States

Forecasting manufacturing defect rates is a crucial process improvement and quality control step. Due to the unavailability of direct data on defect rates, this study uses macro-level data on capacity utilization in the U.S. manufacturing sector as a proxy for defect rate, since process inefficiencies are often signaled by significant deviations in capacity utilization. We construct a novel forecasting framework that compares models built on internal operational variables with models enhanced by external risk indicators, including the economic policy uncertainty index, global supply chain pressure index, labor strikes data, and energy input prices for industrial electricity and natural gas. We form a unique dataset from different sources such as FRED, the New York Federal Reserve, the Bureau of Labor Statistics, the Energy Information Administration, and Economic Policy Uncertainty. We combine the data in a monthly panel structure over multiple years and apply various machine learning models, including autoregressive models, random forests, and gradient boosting with cross-validation to compare out-of-sample accuracy. Our study aims to determine whether external risk indicators, in addition to internal operational variables, improve the forecast accuracy of macro-level process performance and provide a scalable, replicable, and public data-driven framework for quality monitoring. This work also bridges the gap between macroeconomic indicators and forecasting manufacturing inefficiencies, offering a unique approach to data-driven performance analysis.

7 - Uncertainty-Aware Load Planning with Deep Learning for the Parcel Service Industry

Thomas Bruys, Georgia Institute of Technology, Atlanta, GA, United States, Reza Zandehshahvar, Amira Hijazi, Pascal Van Hentenryck

In large-scale logistics operations, load plans, which determine when and where shipments will be processed, are essential for ensuring smooth and efficient deliveries. These plans coordinate the flow of goods through processing centers and help allocate resources. However, with the rise of e-commerce, growing uncertainties like delays and demand shifts often require frequent adjustments to the predefined load plans. This study develops a deep learning-based approach to automate inbound load plan adjustments for a large transportation and logistics company. Leveraging over 250,000 historical load records, the study develops a two-stage framework that incorporates uncertainty quantification using conformal prediction. In the first stage, tactical planning, the system predicts the intended processing building and time window for each load a week before operations. This enables early identification of potential plan adjustments and supports efficient resource allocation. The second stage refines these predictions on the day of operation by incorporating real-time data, such as estimated arrival times, which significantly improves the accuracy and adaptability of the system. This dual-stage process balances long-term preparation with short-term flexibility, a crucial capability in high-volume logistics environments. A key innovation is the integration of conformal prediction, which quantifies uncertainty by producing prediction sets with guaranteed coverage, enhancing trust in model outputs. The deep learning architecture employs dedicated embedding layers for both categorical and numerical features, significantly boosting performance on complex tabular data. Experimental results using an industrial dataset show a high prediction accuracy, with substantial improvements for shifted loads in the second stage.

8 - Resilient Part Allocation for Colocation Data Centers Using Multi-Agent AI: An On-demand Optimization Framework

Neda Zarayeneh, Hitachi America, Santa Clara, CA, United States, Pegah Mavaie, Malarvizhi Sankaranarayanan, Omunshu Thapliyal, Ravigopal Vennelakanti

Colocation data centers face a growing challenge in allocating limited critical equipment—such as UPS systems, PDUs, and generators—across 100+ facilities, each at varying phases of development or operation. Unlike retail inventory systems, data center part allocation is constrained by long lead times (up to 150 weeks), variable grid access timelines, and tightly coupled construction readiness schedules. Inefficient inventory management and misallocation of critical equipment, parts, and resources to expansion or new facility projects can lead to overstocking, missed commissioning milestones, and idle capital. With over 30,000 parts to allocate across facilities, operators must manage inventory decisions with precision. We propose a multi-agent AI-based optimization framework that improves allocation visibility, responsiveness, and accuracy across the data center supply chain. The architecture includes an intelligent agent that automates feature extraction from diverse sources, a dynamic classification layer mapping part criticality to construction milestones, and a set of agents that make collaborative decisions under uncertainty. Our system addresses challenges like part delivery misalignment, reallocation delays, and schedule deviation by incorporating demand signals, delivery buffers, supplier risk scores, and facility readiness in real time. Using scenario simulations and project data from colocation providers, the framework demonstrates approximately 70% reduction in misallocation rate, a 2x improvement in reallocation lead time, and significant gains in fulfillment KPIs such as On-Time, In-Full (OTIF) and Order Cycle Time. Inventory KPIs like Stockout Rate and Inventory Turnover Ratio also improve, reflecting better efficiency and prioritization. This work offers a scalable, adaptive model for resilient infrastructure supply chains.

SE64

Building B Level 4 B404

Panel on Management Science Replication Initiative

Panel Session

Behavioral Operations Management

Co-Chair: Kyle Hyndman, UT Dallas, Richardson, TX, USA, United States

1 - Panelist

Kyle Hyndman, UT Dallas, Richardson, TX, United States

2 - Panelist

Stephen Leider, University of Michigan, Ann Arbor, MI, United States

3 - Panelist

Elena Katok, The University of Texas at Dallas, Richardson, TX, United States

4 - Panelist

Ming Jiang, Shanghai Jiao Tong University, Shanghai, China, People's Republic of

SE65

Building B Level 4 B405

Data-Driven Approaches to Navigate Human-AI Interactions

Invited Session

Artificial Intelligence

Chair: Agni Orfanoudaki, Oxford University, Oxford, N/A, United Kingdom

1 - Learning to Plan Adaptive Human-AI Interaction

Arlen Dean, Washington University in St. Louis, St. Louis, MO, United States, Agni Orfanoudaki, Mohammad Zhalechian

As algorithmic systems are increasingly adopted in the workplace, a key challenge is structuring their use to support, not undermine, human capabilities. This study explores policies for effective human-AI collaboration in a decision-support context. We propose a reinforcement learning (RL) approach to dynamically plan the level of algorithmic guidance provided to a human decision-maker based on their observed user behavior. Our approach aims to navigate the potential risks of both under- and over-reliance, where users may either overlook valuable insights or become overly dependent on them. We theoretically characterize the performance of our proposed model across varying behavioral dynamics and then validate its effectiveness through a controlled experimental study. Our results show that the proposed RL framework can meaningfully improve decision outcomes while sustaining user learning and engagement. We further use our findings to draw insights into when richer behavioral model-derived policies offer added value for planning human-algorithm collaboration.

2 - AI Insurance for Business Operations

Claudio Sánchez Perez de Amézaga, IESE Business School, Barcelona, Spain, Mihalis Markakis, Agni Orfanoudaki

Artificial Intelligence (AI) systems introduce potential risks during implementation and usage when prediction errors lead to costly failures. Insurance, a critical tool for managing risk in many industries, offers a complementary solution to regulatory oversight by providing financial coverage and incentivizing responsible AI deployment. This work proposes a parametric insurance framework to manage such risks in predictive AI applications, providing financial protection to users. We characterize equilibrium contracts that balance user utility, developer incentives, and insurer profitability, accounting for accuracy-sensitive demand and fixed claim costs. Using Inverse Conformal Prediction to model error distributions, we optimize contract terms to maximize developer income. To empirically validate our findings, we leverage a case study on battery state-of-health estimation.

3 - The Value of Partial Information in Sequential Medical Scanning Decisions

Ali Kaan Kurbanzade, Kelley School of Business, Bloomington, IN, United States, Agni Orfanoudaki, Mohammad Zhalechian

In medical decision-making, practitioners often face trade-offs between waiting and conducting diagnostic scans, particularly in areas such as mammography, biopsy, and CT scanning. Existing models typically assume a binary action space—either waiting or performing an expensive full scan. In this study, we extend these models by introducing additional decision options, including partial scans, to assess their impact on sequential decision-making. Partial scans are assumed to be less costly than full scans but also less accurate. However, they provide more reliable information than simply waiting. We build upon existing POMDP frameworks by expanding the action set beyond the traditional binary choices and present theoretical insights into the value of partial information. To validate our findings, we apply our model to a novel dataset from a partnering hospital, focusing on subarachnoid hemorrhage diagnosis and management.

4 - Conformal Decision Risk Assessment

Wenbin Zhou, Carnegie Mellon University, Pittsburgh, PA, United States, Agni Orfanoudaki, Shixiang Woody Zhu

High-stakes decisions in domains such as healthcare, energy, and public policy are often made by human experts using domain knowledge and heuristics, yet are increasingly supported by predictive and optimization-based tools. A dominant approach in operations research is the predict-then-optimize paradigm, where a predictive model estimates uncertain inputs, and an optimization model recommends a decision. However, this approach often lacks interpretability and can fail under distributional uncertainty - particularly when the outcome distribution is multi-modal or complex - leading to brittle or misleading decisions. In this paper, we introduce CREDO, a novel framework that quantifies, for any candidate decision, a distribution-free upper bound on the probability that it is suboptimal. By combining inverse optimization geometry with conformal prediction and generative modeling, CREDO produces finite-sample risk certificates that are both statistically rigorous and practically interpretable. This framework enables human decision-makers to audit and validate their own decisions under uncertainty, bridging the gap between algorithmic tools and real-world judgment.

SE66

Building B Level 4 B406

Learning Outcomes, Creativity, and Generative AI

Panel Session

INFORMS Committee on Teaching and Learning

Co-Chair: Graeme Warren, Johns Hopkins University, Washington, DC, 20001-2114, United States

1 - Panelist

Graeme Warren, Johns Hopkins University, Washington, DC, United States

2 - Panelist

Joseph Cazier, Arizona State University, Tempe, AZ, United States

3 - Panelist

Guiping Hu, Oklahoma State University, Stillwater, OK, United States

4 - Panelist

Ratna Chinnam, Wayne State University, Detroit, MI, United States

5 - Panelist

Parag Dhumal, Arizona State University, Arizona, AZ, United States

SE67

Building B Level 4 B407

Advanced Machine Learning Methods for Data-Driven Analytics on Complex Big Data Structures

Invited Session

Data Mining

Chair: Jaeseung Baek, Northern Michigan University, Marquette, MI, United States

1 - Spatial Adaptive Fused Lasso for Monitoring and Diagnosing Topographic Variations

Jaeseung Baek, Northern Michigan University, Marquette, MI, United States

In this research, we present a new approach for monitoring and diagnosing local topographic variations on the surface topography. Monitoring topographic variations in the surface is crucial for quality engineers since the change in the surface finish is closely related to the performance of products in use. Moreover, surface inspection becomes more difficult when the surface contains local defects that are often concealed due to the high dimensional data structure of the surface image. To tackle this problem, we develop a variable selection-based fault detection and diagnostic approach that selects potentially defective areas on the surface by considering the spatial structure of the surface and residual information. Suspicious areas are then employed to detect anomalies and further utilized to identify defect size, location, and shape for fault diagnosis purposes.

2 - Weighted Adaptive Supervised Multi-Group Canonical Correlation Analysis for Multi-Source Data

Tianhui Wang, Iona University, New Rochelle, NY, United States, Yifan Xie, Myong-Kee Jeong

Integrating multi-source data from diverse equipment and processes is critical for improving the prediction accuracy of key manufacturing parameters, such as critical dimensions, which directly influence the overall quality of the final product. In this context, canonical correlation analysis (CCA) has been widely adopted to enhance the correlation between groups while reducing heterogeneity and noise across data sources. However, existing CCA-based methods focus on modeling inter-source correlations and overlook utilizing output information to enhance predictive performance.

Furthermore, current supervised CCA methods assign equal importance to all group correlations, limiting their capacity to handle low-correlation groups and reducing the predictive accuracy.

To address these challenges, we propose a weighted adaptive supervised multi-group CCA framework that maximizes the sum of correlations between data groups and the output, thereby enhancing the predictive capacity of subsequent models. Adaptive inter-group weighting coefficients are designed to strengthen high-correlation groups while reducing the influence of low-relevance ones, enabling a more focused alignment between each data group and the output.

Experimental results demonstrate that the proposed method significantly strengthens correlations with the output, improves inter-group relationships, and outperforms existing methods in predictive tasks. The method achieves improved computational efficiency when applied to high-dimensional datasets by providing closed-form solutions at each iteration. The proposed method provides an efficient and robust solution for multi-source data integration, particularly in manufacturing processes where quality control and process optimization are critical, while also advancing the theoretical understanding of multi-group CCA.

3 - Multi-Class Support Vector Data Description Fire Monitoring System Using Dynamic TIME Warping Kernel in Different Non-Fire Environments

Omar Alturkistani, Rutgers University, Piscataway, NJ, United States, Taha Alhindi, Jaeseung Baek, Myong Kee Jeong

On a global scale, fire disasters present a serious risk to human well-being and the natural surroundings and infrastructure. Traditional fire detection methods encounter a challenge of data imbalance, with an abundance of non-fire scenario data but limited availability from actual fire incidents, resulting in challenges in building an accurate model capable of detecting various types of fires. Moreover, in real-life situations, sensor data collected from the non-fire scenario may exhibit patterns that fall into more than one class. This specific scenario type, known as nuisance scenarios (e.g., smoking and cooking), shares some characteristics with fire scenarios but is considered non-fire, and neglecting such cases can result in delayed fire detection or trigger nuisance alarms. To address this issue, this paper proposes a novel fire monitoring system based on a multi-class support vector data description with a dynamic time warping kernel, which considers distinct sensor patterns from multiple non-fire scenarios. The developed system is able to effectively detect different fire types without prior knowledge by constructing distinct decision boundaries for different non-fire classes. In addition, the proposed system further considers the temporal dynamic and shape patterns present within sensor signals by incorporating the dynamic time warping kernel function. Experimental results demonstrate the proposed system's superiority over existing methods in detecting different fire types earlier and more accurately with lower false alarm rates in non-fire scenarios.

SE68

Building B Level 4 B408

Responsible, Interpretable, and Trustworthy Machine Learning

Invited Session

Data Mining

Chair: Hadis Anahideh, University of Illinois Chicago, Chicago, IL, United States

1 - Individual-Level Model Selection via Local Rashomon Set Exploration

Parian Haghighat, University of Illinois at Chicago, Chicago, IL, United States, Hadis Anahideh, Cynthia Rudin

In high-stakes domains such as finance and criminal justice, model multiplicity—where many predictive models achieve similar accuracy yet differ in their individual-level decisions—poses serious challenges to fairness, reliability, and interpretability. This phenomenon, known as the Rashomon Effect, creates ambiguity in decision-making and undermines trust in algorithmic systems. In this talk, we present a novel framework for individualized model selection that embraces model multiplicity as an opportunity for context-sensitive and ethically aligned decision-making. Our approach leverages the Rashomon set of sparse decision trees, enumerated using the TreeFARMS algorithm, and introduces a nonlinear manifold learning technique—diffusion maps—to uncover intrinsic data geometry and define personalized neighborhoods. For each individual, we select a locally optimal model based on both local calibration accuracy and adversarial robustness, quantified using Sparse Explanation Values (SEV). Through empirical studies on the FICO credit risk and COMPAS recidivism datasets, we demonstrate that our method significantly improves local accuracy, interpretability, and robustness compared to global and clustering-based baselines.

2 - Multi-Interval Risk-Based Pareto Dominance for Water Network Rehabilitation Under Uncertainty

Raghavendra Punugu, The University of Texas at Arlington, Arlington, TX, United States, Uthman Kareem, Jay Rosenberger, Victoria Chen, Mohsen Shahandashti

In multi-objective optimization, the dominance of a solution is usually determined using point estimates on the objectives. However, this approach may result in incorrectly identifying a solution as non-dominated when in fact it is dominated when considering the uncertainty of the estimators. Although a few studies in the literature have considered incorporating uncertainty in multi-objective optimization using confidence intervals, they have done so with simplifying assumptions or with simple properties on the confidence intervals, such as half-width or midpoint of an interval. To overcome the challenges, we propose an Interval Risk-based Pareto dominance (IRBPD) approach for multiple risk measures such as Value-at-Risk, Conditional Value-at-Risk, and standard deviation. The proposed approach is applied to water distribution network rehabilitation to identify a set of non-dominated rehabilitation strategies to enhance network resilience under seismic uncertainty. The results obtained demonstrate the effectiveness of the proposed approach.

3 - Adaptive Multi-Objective Bayesian Optimization for Fair Hyperparameter Tuning

Nazanin Nezami, University of Illinois Chicago, Chicago, IL, United States, Hadis Anahideh, Mohammadsina Almasi

Hyperparameter tuning plays a crucial role in machine learning (ML) model selection, impacting both predictive accuracy and fairness outcomes. While Multi-Objective Bayesian Optimization (MOBO) provides an effective framework for balancing competing objectives, existing MOBO methods do not explicitly guide the search toward fairness-improving hyperparameters, often leading to inefficient exploration and suboptimal fairness-accuracy trade-offs. In this work, we propose a novel fairness-aware MOBO framework that introduces a Pareto-based coordinate selection strategy to dynamically prioritize fairness-sensitive hyperparameters. Our approach also incorporates a feature importance-driven discretization mechanism to refine the search space, ensuring a more targeted and efficient optimization process. Unlike existing fairness-aware Bayesian optimization methods that either penalize unfair solutions or rely on general-purpose multi-objective search strategies, our method actively steers exploration toward fairness-optimal configurations. We evaluate our approach against state-of-the-art MOBO techniques, including ParEGO and EHVI, demonstrating that our method achieves superior fairness-accuracy trade-offs while maintaining computational efficiency. These results highlight the importance of fairness-aware exploration in hyperparameter tuning and establish our approach as a scalable solution for responsible model selection.

SE69

Building B Level 4 B409

Incentives, Alliances, and Evaluation in NPD

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Zhi Chen, National University of Singapore, Singapore, Singapore

Co-Chair: Guiyun Feng, Singapore Management University, Singapore, Singapore

1 - Sourcing Innovation and Production: a Mechanism Design Approach

Meng Chen, Southern University of Science and Technology, Shenzhen, Guangdong, China, People's Republic of, Xiaoshuai Fan, Ersin Korpeoglu, Cuihong Li

This study investigates the optimal mechanism for a buyer (she) who simultaneously sources innovation design and production from two suppliers (he). The buyer initially commits to a sourcing mechanism that specifies the rules for selecting the design and assigning production. The sourcing mechanism proceeds in two stages. In the first stage, suppliers exert privately known efforts to develop innovation designs with uncertain values. The buyer then provides bonuses based on her report of the innovation values. In the second stage, she selects one design and assigns production to one supplier, with each supplier privately knowing his own production cost. A key feature of our model is that the value of each innovation design is the buyer's private information - a situation commonly observed in many industries where value assessments are inherently subjective and unverifiable by suppliers. This asymmetric information poses significant challenges in designing effective contracts. Our analysis shows that when innovation values are unobservable, bonus payments - typically secondary to allocation adjustments - become essential for ensuring truthful reporting by the buyer.

Moreover, when the effort cost is high, the mechanism can no longer rely solely on monetary incentives, as in the observable case. Instead, it strategically intensifies allocation distortions under low-value innovations, sometimes to the extent that no supplier is selected for production. This research contributes to the literature by illustrating how the privacy of innovation values reshapes the structure of the optimal procurement mechanism.

2 - AI Supervisors and Human Creativity

Sanjiv Erat, University of California-San Diego, La Jolla, CA, United States, Ping-Chieh Huang

With AI-based tools improving at an exponential pace, the day is perhaps not far off when management of knowledge work and of knowledge workers becomes just another skill that can be performed by AI supervisors. The current study investigate how a person's creative performance is affected by the identity of the feedback giver and the nature of the feedback.

3 - Rank-Oriented Benchmarking of Academic Commercialization

Cagri Ozmemis, Georgia Institute of Technology, Atlanta, GA, United States, Karthik Ramachandran

We introduce a rank-oriented benchmarking framework to evaluate academic commercialization performance, reflecting universities' strategic focus on rankings. Built on the objectivity of Data Envelopment Analysis (DEA), our approach prioritizes relative rank over traditional efficiency scores and reveals a counterintuitive trade-off: institutions can attain better ranks under rank-oriented DEA while appearing less efficient than under standard efficiency-oriented DEA. Compared to averaging-based benchmarking method, our approach offers sharper discrimination among peers, a critical feature for rankings. Applying our framework to the AUTM Licensing Survey, we find that universities distributing effort more evenly across multiple commercialization outputs tend to receive more favorable peer-evaluated rankings under rank-oriented evaluation than under efficiency-oriented evaluation. Notably, institutions affiliated with a University Affiliated Research Center (UARC) receive relatively lower evaluations by their peers specifically due to their affiliation.

4 - Sequential Versus Parallel Search Under Competition

Yufei Zhu, National University of Singapore, Singapore, Singapore, Zhi Chen, Zhenyu Hu

Concept generation and new product development (NPD) are essential for business success, as they enable firms to stay innovative in a highly competitive market. However, due to the inherent uncertainty and complexity of the innovation landscape, it is essential that firms thoroughly search across the space of possible concepts, thereby increasing their chances of discovering a winning idea. In practice, sequential and parallel search strategies represent two fundamental approaches to concept generation and NPD. However, firms oftentimes do not conduct concept generation internally but instead delegate it to multiple external agents, such as suppliers or service providers. In this paper, we ask: in such a competitive, multi-agent setting, should the buyer prefer each supplier to conduct a sequential or parallel search? We employ a contest approach to answer this question. Our analysis reveals that when the unit search cost is high or the maximum number of evaluation opportunities is large, the buyer prefers sequential search. Otherwise, parallel search is preferred.

Sunday, October 26, 5:45 PM - 6:35 PM

Building A Level 4 A411

Keynote - How can INFORMS contribute to The Second Quantum Revolution?

Plenary Keynote

Keynote

1 - How Can INFORMS Contribute to the Second Quantum Revolution?

Sridhar Tayur, Carnegie Mellon University, Pittsburgh, PA, United States

The Second Quantum Revolution – comprising of Quantum Computing, Quantum Communications and Quantum Sensing – provides unprecedented abilities to improve the human condition, security and sustainability, hence mass prosperity, across the world, in a variety of ways, ranging from improving health (through better diagnosis as well new drug development, via exciting applications of Quantum Machine Learning and Quantum Sensing) as well by improving security (in financial transactions as well as better combat cyber-attacks on critical infrastructure, through Quantum Money and Quantum Communication protocols). It also promises to do so using less energy than purely digital methods, thus providing a sustainable approach to sustainable development in a hybrid quantum-classical future.

Integer Programming, Queueing, Markov Decision Processes, Semi-definite Programs. These are some of fundamental methodologies in Operations Research and Management Science that are used to tackle practical applications from Business (Supply Chain, Finance), Engineering (Communication Networks) and Medicine (Cancer Genomics, Image Recognition). At Tepper Quantum Group, we are exploring the twin questions: (a) what can quantum do for OR/MS and (b) what can OR/MS do for quantum.

In this talk, after a brief introduction to the relevant topics in quantum mechanics, I will present an overview of our progress in (a) Graver Augmented Multi-seed Algorithm (GAMA), a novel hybrid quantum-classical algorithm, for non-linear integer programs, with application to hedge funds, cancer genomics and supply chains; (b) Ising Hardware (Photonics); (c) Queueing analyses for quantum communications (buffering, entanglement switch) and (d) tackling fundamental problems in Quantum Information Science (QIS), such as entanglement detection, via SDP.

Building B Level 2 B206

IFORS Distinguished Lecture

Plenary Keynote

Keynote

1 - 40 Years of Multi-Objective Evolutionary Algorithms: past, Present and Future

Carlos Coella Coella, CINVESTAV-IPN, Mexico City, Mexico

The first multi-objective evolutionary algorithm was proposed in 1985 by David Schaffer (in the USA). Since then, this research area has become increasingly popular and has generated an important number of publications and PhD theses. In this talk, I'll give a quick overview of the historical development of this discipline, including some of the recent research topics that are attracting interest as well as some of the areas in which more research is expected in the future.

Building A Level 4 A412

Keynote- Optimization, ML and AI in Operations Management

Plenary Keynote

Keynote

1 - Optimization, ML and AI in Operations Management

Georgia Perakis, Massachusetts Institute of Technology, Cambridge, MA, United States

Data-driven decision-making has garnered a growing interest due to the increase in data availability in recent years. The growth of AI has further accelerated this growth. With that growth, many opportunities as well as challenges arise. In this talk, we will discuss some of these opportunities and challenges as well as many applications in Operations Management. We will highlight the importance and challenges of integrating optimization with ML in data-driven decision-making as well some concrete examples of these synergies. For example, we will discuss how we can optimize over already trained objective functions that arise either from tree ensemble predictive models or from neural network models in order to recommend better decisions.

Sunday, October 26, 5:45 PM - 6:45 PM

Building A Level 4 A411

Keynote - How can INFORMS contribute to The Second Quantum Revolution?

Plenary Keynote

Keynote

1 - How Can INFORMS Contribute to the Second Quantum Revolution?**Sridhar Tayur, Carnegie Mellon University, Pittsburgh, PA, United States**

The Second Quantum Revolution – comprising of Quantum Computing, Quantum Communications and Quantum Sensing – provides unprecedented abilities to improve the human condition, security and sustainability, hence mass prosperity, across the world, in a variety of ways, ranging from improving health (through better diagnosis as well new drug development, via exciting applications of Quantum Machine Learning and Quantum Sensing) as well by improving security (in financial transactions as well as better combat cyber-attacks on critical infrastructure, through Quantum Money and Quantum Communication protocols). It also promises to do so using less energy than purely digital methods, thus providing a sustainable approach to sustainable development in a hybrid quantum-classical future.

Integer Programming, Queueing, Markov Decision Processes, Semi-definite Programs. These are some of fundamental methodologies in Operations Research and Management Science that are used to tackle practical applications from Business (Supply Chain, Finance), Engineering (Communication Networks) and Medicine (Cancer Genomics, Image Recognition). At Tepper Quantum Group, we are exploring the twin questions: (a) what can quantum do for OR/MS and (b) what can OR/MS do for quantum.

In this talk, after a brief introduction to the relevant topics in quantum mechanics, I will present an overview of our progress in (a) Graver Augmented Multi-seed Algorithm (GAMA), a novel hybrid quantum-classical algorithm, for non-linear integer programs, with application to hedge funds, cancer genomics and supply chains; (b) Ising Hardware (Photonics); (c) Queueing analyses for quantum communications (buffering, entanglement switch) and (d) tackling fundamental problems in Quantum Information Science (QIS), such as entanglement detection, via SDP.

Building B Level 2 B206

IFORS Distinguished Lecture

Plenary Keynote

Keynote

1 - 40 Years of Multi-Objective Evolutionary Algorithms: past, Present and Future**Carlos Coella Coella, CINVESTAV-IPN, Mexico City, Mexico**

The first multi-objective evolutionary algorithm was proposed in 1985 by David Schaffer (in the USA). Since then, this research area has become increasingly popular and has generated an important number of publications and PhD theses. In this talk, I'll give a quick overview of the historical development of this discipline, including some of the recent research topics that are attracting interest as well as some of the areas in which more research is expected in the future.

Building A Level 4 A412

Keynote- Optimization, ML and AI in Operations Management

Plenary Keynote

Keynote

1 - Optimization, ML and AI in Operations Management**Georgia Perakis, Massachusetts Institute of Technology, Cambridge, MA, United States**

Data-driven decision-making has garnered a growing interest due to the increase in data availability in recent years. The growth of AI has further accelerated this growth. With that growth, many opportunities as well as challenges arise. In this talk, we will discuss some of these opportunities and challenges as well as many applications in Operations Management. We will highlight the importance and challenges of integrating optimization with ML in data-driven decision-making as well some concrete examples of these synergies. For example, we will discuss how we can optimize over already trained objective functions that arise either from tree ensemble predictive models or from neural network models in order to recommend better decisions.

Monday, October 27, 8:00 AM - 9:15 AM

MA01

Building A Level 3 A301

Provalis Research/Artelys

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Beyond the Hype: Rethinking Text Analysis in the Age of Generative AI**Normand Peladeau, Provalis Research, Montreal, QC, Canada**

In this presentation, we will explore latest advancements in generative AI and large language models (LLMs) and see their contribution as well as limitations for text analytics applications. While the excitement surrounding such new technologies is undeniable, older techniques still provide significant value and may even outperform newer methods in certain contexts. By critically examining the strengths and limitations of both established and cutting-edge approaches, we want to offer a balanced, realistic perspective on the role of recent developments in NLP and AI.

2 - Model and Solve Nonlinear Optimization Problems with Artelys Knitro**Richard Waltz, Artelys, Los Angeles, CA, United States**

Artelys Knitro is the leading solver focused on large-scale, nonlinear (potentially non-convex), optimization problems. Knitro offers both interior-point and active-set algorithms for continuous models, as well as tools for handling problems with integer variables and other discrete structures. This tutorial will introduce the key features of Knitro, and demonstrate how to use Knitro to model and solve optimization problems in various environments. The talk will showcase several instances of nonlinear problems, typically addressed through linear relaxation, yet warranting a direct approach. It will delve into the methodologies and tools employed to tackle these challenges. This tutorial aims at an audience familiar with the basics of mathematical optimization, focusing on practical examples.

MA02

Building A Level 3 A302

K-12 Outreach: Collaboration across INFORMS Societies

Panel Session

Committee's Choice

Co-Chair: Yifeng Wang, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Zihan Zhang, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Fenglian Pan, The University of North Carolina at Charlotte, Charlotte, NC, United States

1 - Moderator Panelist

Fenglian Pan, University of Arizona, Tucson, AZ, United States

2 - Moderator Panelist

Yifeng Wang, Georgia Institute of Technology, Atlanta, GA, United States

3 - Moderator Panelist

Zihan Zhang, Georgia Institute of Technology, Atlanta, GA, United States

4 - Panelist

Laura Albert, University of Wisconsin-Madison, Madison, WI, United States

5 - Panelist

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States

6 - Panelist

Ken Murphy, UC Irvine, Irvine, CA, United States

7 - Panelist

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

MA03

Building A Level 3 A303

Advances in Multiobjective Optimization

Invited Session

Multi Criteria Decision Making

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

Co-Chair: Kristen Joyce, Florida Polytechnic University, Lakeland, FL, 33896, United States

1 - Pareto Sensitivity, Most-Changing Sub-Fronts, and Knee Solutions

Tommaso Giovannelli, University of Cincinnati, Cincinnati, OH, United States, Marcos Medeiros Raimundo, Luis Vicente

When dealing with a multi-objective optimization problem, obtaining a comprehensive representation of the Pareto front can be computationally expensive. Furthermore, identifying the most representative Pareto solutions can be difficult and sometimes ambiguous. A popular selection are the so-called Pareto knee solutions, where a small improvement in any objective leads to a large deterioration in at least one other objective. In this paper, using Pareto sensitivity, we show how to compute Pareto knee solutions according to their verbal definition of least maximal change. We refer to the resulting approach as the sensitivity knee (snee) approach, and we apply it to unconstrained and constrained problems. Pareto sensitivity can also be used to compute the most-changing Pareto sub-fronts around a Pareto solution, where the points are distributed along directions of maximum change, which could be of interest in a decision-making process if one is willing to explore solutions around a current one. Our approach is still restricted to scalarized methods, in particular to the weighted-sum or epsilon-constrained methods, and require the computation or approximations of first- and second-order derivatives. We include numerical results from synthetic problems that illustrate the benefits of our approach.

2 - Set-Based Regret and Algorithms for Online Multiobjective Optimization

Kristen Joyce, Florida Polytechnic University, Lakeland, FL, United States, Margaret Wiecek

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

Scalar regret and set-based regret are defined for OMO and used to evaluate the performance of the Online Gradient Descent (OGD) Algorithm. Minimizing the scalarized OMO problem, the OGD computes a single solution in every iteration and yields the scalar regret. Minimizing the generational distance or inverted generational distance associated with the OMO problem, the OGD computes a set of solutions at every iteration and yields the set-based regret. Numerical results are presented that highlight the value of set-based algorithms over single-solution algorithms corresponding to scalarized OMO problems. Additionally, an online biobjective application is presented for an autonomous vehicle driving on unknown terrain that minimizes both time elapsed and energy loss.

3 - Synthetic Tradespaces for Tradespace Optimization

Cameron Turner, Clemson University, Clemson, SC, United States

A significant challenge for tradespace optimization is the lack of real world tradespace data. However, through the use of synthetic data, it becomes possible to produce representative tradespace data sets without the need for expensive or proprietary data enabling the study and evaluation of tradespace optimization techniques. Eliminating the need to acquire real tradespace data is a significant accomplishment as the costs associated with acquiring real tradespace data and then curating that data are significant in terms of time, effort and money. In this presentation, a sample application for the generation of synthetic tradespace data is introduced along with the necessary techniques to populate the tradespace with a dataset. This dataset can be explored, and analyzed to reveal interesting applications for the tradespace studies. Ultimately, the tradespace can be subjected to an optimization process to identify solutions to the tradespace optimization problem. The sample problem presented has been used in an academic setting for the last five years to allow students to wrestle with the challenges of optimizing in highly complex, multidimensional, nonlinear, mixed-continuous/discrete tradespaces in a realistic manner, while remaining cost-effective to maintain and update to remain technically current and plausible.

4 - Equity-Driven Workload Allocation for Crowdsourced Last-Mile Delivery: a Biobjective Optimization Approach

Hadi Gard (Charkhgard), University of South Florida, Tampa, FL, United States, Iman Dayarian, Abhay Sobhanan

Crowdshipping, a rapidly growing approach in Last-Mile Delivery (LMD), relies on independent crowdworkers for delivery orders. Building a sustainable network of crowdshippers is essential for the survival and growth of such systems, while their participation is primarily motivated by fair pay. Additionally, the financial well-being of crowdworkers is sensitive to fair compensation, especially for those who depend on crowdwork as their main source of income. Therefore, equitable workload allocation and compensation mechanisms in crowdsourcing platforms will benefit both platforms and crowdworkers. We aim to answer several questions gig-economy platforms interested in fair pay may ask: How to measure equity, assess the cost benefits, and manage potential drawbacks? Our main contribution is the proposal of a practical equity-oriented framework tailored to crowdshipping within an LMD environment. This framework draws inspiration from the real-world operations of a group of crowdshipping platforms and operates in real-time. At its core is a bi-objective optimization process that balances equity and cost, aiming to address the study's main research questions. Built on a theoretical foundation, it enables the use of various equity measures and allows us to identify the equity measure that most reliably explores the trade-offs between cost and equity. We show that even a marginal sacrifice in cost efficiency (e.g., 2.5%) can significantly improve equity, potentially up to 39%. We provide actionable recommendations for practitioners, offering insights into selecting equity measures and workforce management.

5 - On the Nondominated Envelope for Multiobjective Optimization

Margaret Wiecek, Clemson University, Clemson, SC, United States, Herve Kerivin, Rakhi Goswami

Multiobjective optimization problems (MOPs) with parametric right-hand-side coefficients model decision-making problems in the presence of multiple, conflicting objective functions, and under uncertainty. For a parametrized family of general MOPs with a parameter drawn from a convex polytope serving as an uncertainty set, we

introduce the concept of nondominated envelope and examine its properties. Specific results are obtained for the linear case.

6 - A Paradigm Shift in Solving Multi-Objective Integer Programs via Objective Domain Reduction

Lagnajita Basu, University of South Florida, Tampa, FL, United States, Hadi Charkhgard

We propose a paradigm shift in generating the Pareto-optimal frontier for multi-objective pure integer programs by introducing binary-encoding reformulations and systematically reducing the objective function value domains. This leads to a new class of branch-and-bound algorithms, where each node corresponds to a feasibility problem solvable via constraint programming. To demonstrate the effectiveness of the proposed method, we focus on bi-objective optimization problems and provide numerical insights and comparisons across a benchmark set of pure integer programs.

MA04

Building A Level 3 A304

Financial Contagion and Risk

Invited Session

Finance

Chair: Çağın Ararat, Bilkent University, Ankara, Turkey

1 - Can Nash Inform Capital Requirements? Allocating Systemic Risk Measures

Zachary Feinstein, Stevens Institute of Technology, Hoboken, NJ, United States, Cagin Ararat

Systemic risk measures aggregate the risks from multiple financial institutions to find system-wide capital requirements. Though much attention has been given to assessing the level of systemic risk, less has been given to allocating that risk to the constituent institutions. Within this talk, we propose a Nash allocation rule that is inspired by game theory. Intuitively, to construct these capital allocations, the banks compete in a game to reduce their own capital requirements while, simultaneously, maintaining system-level acceptability. Sufficient conditions for existence and uniqueness of these Nash allocations are provided with applications provided for the prominent structures used for systemic risk measures in the literature. We demonstrate the efficacy of these Nash allocations with numerical case studies.

2 - Systemic Robustness: a Mean-Field Particle System Approach

Wenpin Tang, Columbia University, New York, NY, United States

This work is concerned with the problem of budget control in a large particle system modeled by stochastic differential equations involving hitting times, which arises from considerations of systemic risk in a regional financial network. Motivated by Tang and Tsai (Ann. Probab., 46 (2018), pp. 1597–1650), we focus on the number or proportion of surviving entities that never default to measure the systemic robustness. First we show that both the mean-field particle system and its limit McKean-Vlasov equation are well-posed by virtue of the notion of minimal solutions. We then establish a connection between the proportion of surviving entities in the large particle system and the probability of default in the McKean-Vlasov equation as the size of the interacting particle system N tends to infinity. Finally, we study the asymptotic efficiency of budget control in different economy regimes: the expected number of surviving entities is of constant order in a negative economy; it is of order root N in a neutral economy; and it is of order N in a positive economy where the budget's effect is negligible.

3 - Systemic Risk Measures: Convexity and Beyond

Çağın Ararat, Bilkent University, Ankara, Turkey

In the literature, systemic risk measures have been introduced as multivariate functionals that calculate acceptable capital allocations for an interconnected financial network. The convexity properties of these functionals are determined both by the aggregation mechanism of the network as well as the acceptability rule for the aggregate quantities. In this talk, we present a duality approach to calculate minimum capital allocations for financial networks under concave aggregation mechanisms and convex acceptance sets. Then, as an example of a non-convex systemic risk measure, we consider the case where the acceptance set is that of value-at-risk. In this case, we quantify the lack of convexity for the set values of the systemic risk measure and provide an upper bound on it.

4 - Dynamics of Cascading Losses in Stochastic Networks

Hamed Amini, University of Florida, Newberry, FL, United States

We study risk processes and cascading defaults in large financial networks that exhibit a locally tree-like structure in the limit. Our analysis extends the classical threshold contagion model by incorporating capital growth, allowing us to examine how a firm's default influences its neighbors and potentially triggers a network-wide cascade. We consider a range of network models, including random regular graphs and general continuous-type inhomogeneous random graphs, to explore the interplay between network topology and risk propagation. We also assess the network's resilience to the initial default of a single node, extending existing resilience frameworks to accommodate dynamic thresholds with capital accumulation.

5 - Incomplete Market Equilibrium with Relative Consumption Criterion: a Mean-Field Game Approach

Marko Hans Weber, National University of Singapore, Singapore, Singapore

We study the general equilibrium of a pure-exchange economy with a continuum of agents who face uninsurable income risk, trade dividend-paying stocks, and engage in mutual borrowing and lending. Agents hold "keeping up with the Joneses" preferences and differ in their risk aversion, time preferences, and relative consumption concerns. Each agent's utility depends not only on their own consumption but also on a weighted average of others' consumption, captured through a weighting kernel. We formulate the equilibrium as a mean-field game and derive a closed-form solution. Finally, we show that the resulting mean-field market equilibrium arises as the limit of equilibria in finite-agent economies as the number of agents tends to infinity.

MA05

Building A Level 3 A305

NSF Proposal Writing Workshop

Invited Session

NSF

Chair: Tapas Das, University of South Florida, Tampa, FL, United States

1 - soon

Tapas Das, University of South Florida, Tampa, FL, United States

tba

2 - test

MA06

Building A Level 3 A306

Funding Opportunities for Data Science Research

Panel Session

Committee's Choice

Co-Chair: Ying Lin, University of Houston, Houston, TX, United States

Co-Chair: Mohammad Bisheh, Georgia Institute of Technology, Atlanta, GA, United States

1 - Moderator Panelist

Mohammad Bisheh, Georgia Institute of Technology, Atlanta, GA, United States

This panel session aims to provide INFORMS members with critical insights into navigating the competitive landscape of data science research funding. Panelists, including representatives from major federal funding agencies like the National Science Foundation (NSF) and the National Institute of Standards and Technology (NIST), alongside experienced academic researchers, will discuss current funding priorities, key program initiatives, and anticipated future directions at the intersection of data science, operations research, statistics, and industrial engineering. The session will cover practical advice on crafting compelling proposals, understanding review processes, and identifying relevant funding opportunities across different agencies. Attendees will gain valuable perspectives on how to align their research agendas with national priorities and increase their success rate in securing grants.

2 - Moderator Panelist

Ying Lin, University of Houston, Houston, TX, United States

This panel session aims to provide INFORMS members with critical insights into navigating the competitive landscape of data science research funding. Panelists, including representatives from major federal funding agencies like the National Science Foundation (NSF) and the National Institute of Standards and Technology (NIST), alongside experienced academic researchers, will discuss current funding priorities, key program initiatives, and anticipated future directions at the intersection of data science, operations research, statistics, and industrial engineering. The session will cover practical advice on crafting compelling proposals, understanding review processes, and identifying relevant funding opportunities across different agencies. Attendees will gain valuable perspectives on how to align their research agendas with national priorities and increase their success rate in securing grants.

3 - Panelist

Satish Bukkapatnam, National Science Foundation, College Station, TX, United States

4 - Panelist

Saman Alani-Azar, Walmart Global Tech, Dallas, TX, United States

5 - Panelist

Rajesh Gupta, Ford Motor Company, Dearborn, MI, United States

6 - Panelist

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

MA07

Building A Level 3 A307

Optimization for Transportation Applications

Job Market Showcase

Job Market Showcase

Chair: Myungeun Eom, Georgia Institute of Technology, Atlanta, GA, United States

1 - Flow Optimization and Coordination for Self-Driving Vehicles in Urban Street Networks

Xiaochen Shi, ISyE Georgia Tech, Atlanta, GA, United States, Anton Kleywegt

Infrastructure guided self-driving (IGSD) technology raises the potential to control the path and speed of IGSD vehicles in such a way that traffic flow efficiency is greatly improved. IGSD vehicles approaching the same intersection are grouped into batches based on non-conflict phases, each of which traverses the intersection without stopping. The greater challenge is to factor into the knock-on effect among intersections, and to create a coordinated solution for IGSD vehicle flow to maximize network performance. We propose a Network Batch Algorithm for 100% IGSD street networks, which determines the velocities of vehicles in such a way that no vehicle has to stop while traversing the network, and that minimizes the total delay. An Access-control (AC) method is also developed to reduce the knock-on effect and consequent gridlock, which can be applied with most traffic controls with minor modification. The approaches are experimented in a dynamic setting with the open-source simulator SUMO for Sioux Falls network and compared with benchmark traffic management schemes. The numerical results demonstrate that Batch algorithm with AC method significantly outperforms other traffic management schemes and increases the throughput of the network by 45%.

2 - Urban Delivery with a Coordinated Fleet of Trucks and Drones

Wenjia ZENG, The Hong Kong University of Science and Technology, Hong Kong, China, People's Republic of, Hai Yang, Hai Wang, MINGYAO QI

With the rapid growth of urban delivery demand due to the fast expansion of e-commerce and urban logistics, the recent development and popularization of unmanned aerial vehicles--drones--have attracted increasing attention from large retailers and logistics companies that seek to use drones in urban delivery. In this paper, we study an urban delivery system with a coordinated fleet of trucks and drones, in which multiple trucks deliver parcels and are resupplied by multiple drones at meeting points. We formulate the coordinated delivery problem as a mixed-integer programming model and propose a novel joint assigning, meeting, and scheduling (JAMES) framework that efficiently yields exact solutions when complete information is available. To the best of our knowledge, this is the first exact method to solve the delivery problem with multiple trucks resupplied by multiple drones. We introduce a novel dynamic programming algorithm embedded in an optimal strategy for time scheduling to accelerate the solution. To accommodate an online environment with dynamically arriving orders, an online policy is deployed to make recurrent decisions. Computational experiments show that the JAMES framework significantly outperforms a commercial solver in computational time, solution quality, and scale. The case study on the dynamic framework demonstrates that the online policy generally outperforms traditional greedy heuristics. Evaluation results for a real-world data set demonstrate the benefits of a coordinated fleet, which are particularly evident when transportation resources are limited, because it can increase drone usage and improve the overall service rate.

3 - Dynamic Non-Bipartite Matching in Transportation

Myungeun Eom, Georgia Institute of Technology, Atlanta, GA, United States, Alejandro Toriello

We study a dynamic non-bipartite stochastic matching problem, where nodes appear following a type-specific independent distribution and wait in the system for a given sojourn time. This problem is motivated by applications in ride-sharing and freight transportation marketplaces, and is related to other on-demand marketplaces. We study the asymptotic properties of two widely used policies, batching and greedy, by analyzing a single-pair case and then converting to the general counterpart using a fluid relaxation and randomization. Finally, we present a computational study simulating freight transportation and ride-sharing marketplaces to assess the empirical effectiveness of the policies. We show that the batching policy is asymptotically optimal with respect to the sojourn time; similarly, while a straightforward greedy policy may not be optimal, a greedy policy with randomized modifications is asymptotically optimal. Perhaps more practically relevant, both policies converge exponentially fast to approximate optimality. We also extend our model to an impatient setting in which each unmatched node leaves at the end of each period with a type-dependent probability. We show that the results for the two policies still hold under different assumptions about the nodes' patience; roughly speaking, the batching policy requires more patient nodes than the greedy policy to remain optimal. Our results suggest that managers can achieve near-optimal performance by using simple greedy or batching policies, with only a reasonably small maximum waiting time guarantee, and even in the presence of potentially impatient nodes.

MA08

Building A Level 3 A310

Healthcare, IT, and Smarter Decision-Making

Invited Session

Information Systems

Chair: Shujing Sun, University of Texas at Dallas, Richardson, TX, United States

1 - Pleasure OR Peril? the Rise of Online spORts Gambling and Its Mental Health Impact**Jiayuan Tian, University of Florida, Gainesville, FL, United States, Guohou Shan, Michael Rivera, Liangfei Qiu, Md Mahmudul Hasan**

The rapid growth of online sports gambling following its legalization in 2018 has outpaced scholarly attention, fueling ongoing public debate about its broader consequences. Although online sports gambling can serve as a form of entertainment—bringing joy and temporary mood boosts, which we refer to as the pleasure effect—it could also intensify gambling disorders that negatively affect mental health, known as the disorder pathway effect. Our study addresses this ambiguity by exploiting a natural experiment and empirically investigating the impact of online sports gambling on mental health using a unique nationally representative panel dataset. Our findings show that online sports gambling deteriorates the mental health of local residents by approximately 3% on average, primarily by exacerbating gambling disorders. We further examine the heterogeneous treatment effects, uncovering a pronounced disorder pathway in rural areas among individuals with existing addiction. At the same time, we identify a weaker disorder pathway effect in regions with higher education levels and better access to primary healthcare. These findings indicate the uneven impacts of online sports gambling due to social and economic factors. Our study has practical implications for both digital gambling platform holders and policymakers to build sustainable online gambling platforms through targeted regulations that mitigate harm. This study complements the IS literature by addressing the underexplored domain of online sports gambling, an emerging Internet-enabled service. From a theoretical perspective, it also contributes to studies on behavioral health economics and digital gambling ecosystems by uncovering causal pathways that link online sports gambling to mental health outcomes.

2 - The Unintended Consequences of Quota-Based Policy on Content Moderation Efficiency: Empirical Evidence from Online Health Platform**Chen Jie, Arizona State University, Tempe, AZ, United States, Kan Xu, Tongxin Zhou, Pei-yu Chen**

Online Health Question-Answer Platforms (OHQPs) provide accessible and timely health information, improving health literacy and promoting equity. However, these platforms face significant moderation challenges due to limited human resources, high inquiry volumes, and the risks of misinformation. To address these issues, some OHQPs have introduced quota-based moderation policies that limit the number of user inquiries. While prior studies suggest quotas can alleviate congestion in other domains, their effectiveness in health information settings remains unclear. This study evaluates a quota-based policy implemented on a leading OHQP focused on HIV-related concerns. Leveraging a natural experiment and constructing a month-level panel dataset spanning one year before and after the policy, we compare a treated English forum with a structurally similar Spanish forum unaffected by the quota. Difference-in-differences analysis reveals that the policy reduced question volume by 67.4% and responses by 68.0%, accompanied by a 13.5% decline in the response rate and no improvement in response time—indicating no gains in moderation efficiency. Textual analysis further reveals behavioral mismatches: moderators increasingly emphasized community rules and politeness marker (e.g., “please,” “thank you”) while referencing fewer health-related topics (e.g., “illness, sexual and users submitted longer, more socially oriented inquiries with greater use of pronouns like “you” and “they,” suggesting attempts to demonstrate community awareness. This study contributes to the literature by providing one of the first empirical evaluations of quota-based moderation in online health platforms, revealing how scarcity policies can unintentionally degrade platform efficiency by misaligning user and moderator behaviors.

3 - How Consumer-Facing Digital Technologies Shape Productivity in High-Stakes Services: Evidence from Kidney Transplantation**Shujing Sun, University of Texas at Dallas, Richardson, TX, United States, Mehmet Ayyaci, Yeongin Kim, Srinivasan Raghunathan, Bekir Tanriover**

Consumer-facing digital technologies allow service users to access information, interact with providers, and participate in decisions. While such technologies have improved service delivery in industries like retail and finance, their role in healthcare is less clear. Healthcare features greater information gaps and shared responsibility between patients and providers, which complicates how and when technology creates value. Motivated by this uncertainty, we study the effect of patient portals—digital tools that provide patients with access to health records and secure messaging—on service efficiency and patient outcomes in kidney transplantation. This setting involves long waits, limited supply, and time-sensitive decisions, making it well-suited for evaluating the value of patient-facing digital technologies. Using national data on transplant patients linked with hospital-level information on portal availability, we examine the impact of portal access on wait time, health status at transplant, and survival after transplant. We find that portal access shortens wait time and improves patient health at the time of transplant. Features that support communication and information access, rather than administrative tasks, are the main drivers of these benefits. The effects are larger for patients in areas with better internet coverage and lower social vulnerability. We also show that improved health at transplant helps explain why portal users have lower post-transplant mortality. Our findings demonstrate how patient-facing technologies can improve process and clinical outcomes in complex care settings, while also highlighting how digital access and social context shape their effectiveness.

4 - Healthcare Under New Ownership: Does Private Equity Improve Healthcare Value?**Minghong Yuan, The University of Texas at Austin, Austin, TX, United States, Wen Wen, Indranil Bardhan**

In recent years, the healthcare sector has experienced a significant increase in private equity (PE) investment. However, the overall impact of such investment on healthcare performance remains insufficiently understood. This study examines the impact of PE investment on healthcare value, focusing on the complex interplay between operational efficiency and patient care quality. While PE firms often prioritize cost-cutting strategies and operational efficiencies to enhance profitability, these financial objectives may conflict with the core mission of healthcare: delivering high-quality and patient-centered care. Using a difference-in-differences approach, we analyze data from U.S. hospitals, assessing changes in healthcare value, defined as the balance between clinical resource use and patient outcomes, following PE investment. We find that overall healthcare value declines after PE investment. We also examine the moderating role of IT-enabled health information sharing and find that robust information-sharing capabilities can mitigate the negative effects of PE ownership. Specifically, hospitals with strong information sharing experience greater reductions in operating costs and improvements in care quality, resulting in higher post-investment care value. Further analysis reveals heterogeneity in the effect of information sharing based on provider type. By distinguishing between hospital-to-hospital and hospital-to-ambulatory provider information sharing, we find that improvements in care quality are primarily driven by enhanced data exchange with ambulatory providers. These results highlight critical need to align investors interest with patient outcomes and point to the importance of supportive policies that encourage investment in health information sharing to preserve healthcare value under PE ownership.

MA09

Building A Level 3 A311

Responsible Artificial Intelligence

Job Market Showcase

Job Market Showcase

Chair: Tongnian Wang, The University of Texas at San Antonio, San Antonio, TX, United States

1 - Exploring the Construction of Local GenAI Governance: a Case of Higher Education Professionals**Rui Sun, New York University, New York, NY, United States**

Generative artificial intelligence (genAI) technologies pose new challenges to AI governance in knowledge work. Aspects of genAI adoption render some traditional governance strategies ineffective. Risks are recognized but difficult to monitor or control. How do knowledge professionals deal with this new reality? How do they construct and mitigate genAI-related risks? This study investigates these questions in the context of higher education, whose core functions of teaching and learning are immensely impacted by students' adoption of genAI tools. I conducted a multi-year inductive qualitative field study of a higher education organization since the emergence of ChatGPT in 2022, focusing on how instructors construct acceptable and unacceptable student AI use in class, and how their local AI governance strategies emerge and evolve through time. Drawing on the social construction of risk object framework, I explicate how various genAI adoption practices in teaching and learning can be constructed as posing harm to the work of higher education professionals. I also illustrate how the professionals try to mitigate the perceived genAI risks by

constructing different local policies and work adaptations. This study presents an effort to explore the emerging practices of genAI governance in knowledge work. It contributes to research on AI governance by showing how aspects of technology adoption and aspects of knowledge work influence the local governance of AI risks.

2 - When Systemic Biases Taint Algorithms: A Path to More Equitable Access in Healthcare

Ozgur Aksoy, The University of Texas at Dallas, Richardson, TX, United States, Mehmet Ayvaci, Asunur Cezar, Srinivasan Raghunathan

Predictive algorithms hold promise for improving resource allocation decisions in socially critical contexts, yet they often exhibit algorithmic bias, inadvertently favoring the socially advantaged. Existing mitigation strategies typically focus solely on correcting predictions, neglecting downstream effects on subsequent decisions and the societal disparities that perpetuate this bias. We investigate the effects of algorithmic bias arising from societal disparities within an integrated prediction–decision framework. Specifically, we assess how explicitly incorporating social factors into prediction–decision pairs influences outcomes, particularly in the context of a healthcare payer’s decision to allocate care management programs to beneficiaries at risk of high costs. We develop an analytical and empirical model of the payer’s allocation decision, and compare outcomes when predictions are adjusted using social factors versus when they are not. While such adjustments can mitigate the negative impact of algorithmic bias at the prediction stage, reducing bias does not necessarily mitigate disparities in decision outcomes—especially when intervention effectiveness varies significantly across social groups. Moreover, tight budget constraints on intervention can diminish the value of adjustments for outcome disparity reduction. Using Medicare insurance payments and CDC Social Vulnerability Index data, we develop a predictive algorithm, quantify algorithmic bias, and demonstrate that our method enhances fairness—intervening on 16.56% more disadvantaged beneficiaries—while preserving decision efficiency. Although we focus on health insurance, our approach is applicable to other socially and economically significant domains within and beyond healthcare.

3 - Responsible AI in Critical Care: Toward Trustworthy Clinical Decision-Making

Tongnian Wang, The University of Texas at San Antonio, San Antonio, TX, United States

Artificial intelligence (AI) is transforming clinical decision-making with the growing use of electronic health records (EHRs). However, ensuring outcomes that are both equitable and accurate remains a critical challenge. Biases embedded in EHR data, as well as those introduced during machine learning (ML) model development, can lead to algorithmic bias, resulting in outcomes that disproportionately harm disadvantaged groups. Efforts to mitigate these biases require diverse and representative datasets, yet strict privacy regulations and data-sharing restrictions make this extremely challenging. Federated learning (FL) offers a privacy-preserving solution by enabling healthcare organizations to collaboratively train ML models without sharing sensitive patient data. However, FL introduces additional challenges in mitigating bias and ensuring robust model performance across diverse clinical settings. Bridging these gaps is crucial to align with the ethical principles of maximizing benefit and minimizing harm. To tackle these issues, we introduce FairPFL, a fairness-aware personalized FL framework designed for clinical prediction with EHRs. Our experiments on real-world EHR datasets demonstrate that FairPFL significantly reduces algorithmic bias while maintaining superior predictive performance, outperforming baseline methods in fairness and performance.

4 - A Novel Fairness-Aware Federated Learning Method in Algorithmic Hiring

zejian wu, University of Utah, SLC, UT, United States, Paul Hu, Da Xu

Algorithmic hiring systems are increasingly adopted in organizational decision-making processes, ranging from sourcing algorithms to identify potential applicants for job openings, selecting resumes for initial screening, and ultimately choosing the final candidate. These systems present significant challenges in ensuring fairness across diverse demographic groups while maintaining privacy and addressing data heterogeneity across various hiring platforms. This study introduces a novel fairness-aware federated clustering method with control variates (FAFC-CV) to address these challenges during candidate selection. By incorporating a demographic parity term into client-specific loss functions and leveraging dynamic clustering, our method effectively mitigates fairness disparities while accommodating the unique characteristics of data from diverse hiring platforms. Theoretically, the proposed method ensures fairness improvements with mathematical guarantees across client groups. Extensive empirical evaluations on real-world hiring datasets demonstrate significant fairness enhancements compared to existing algorithms, with comparable or superior predictive accuracy. This research advances the development of ethical AI by offering a scalable framework for fair and effective algorithmic hiring.

MA10

Building A Level 3 A312

OR for Social Good

Job Market Showcase

Job Market Showcase

Chair: Xiaowei Guo, Clemson University, CLEMSON, United States

1 - Rural School Bus Routing and Scheduling

Prabhat Hegde, Dartmouth College, Hanover, NH, United States, Vikrant Vaze

Long school bus rides adversely affect student performance and well-being. Rural school bus rides are particularly long, incentivizing parents to drive their children to school rather than to opt for the school bus. This in turn exacerbates the traffic congestion around schools, further compounding the problem of long bus rides, creating a vicious cycle. It also results in underutilized school buses and higher bus operating costs per rider. To address these challenges, this paper focuses on the design of rural school bus routes and schedules, a particularly challenging problem due to its unique operational complexities, including mixed loading and irregular road networks. We formalize a rural school bus routing and scheduling model that tackles these complexities while minimizing the total bus ride time of students. We develop an original road network-aware cluster-then-route heuristic that leverages our problem formulation to produce high-quality solutions. For real-world case studies, our approach outperforms status quo solutions by reducing the bus ride times of students by 37-39 %. Furthermore, our iterative routing and scheduling approach can improve school bus use by 17-19 % and reduce car trips that induce congestion near schools by 12-17 %. Many rural school districts share the operational characteristics modeled in this study, suggesting the broad applicability of our approach. Ultimately, by reducing student travel times, increasing school bus utilization, and alleviating congestion near schools, our approach enables rural school district planners to address transportation-related barriers to student performance and well-being.

2 - Enhancing Supply Chain Decisions to Increase Reliable and Affordable Water Access by Adapting the Inventory Routing Problem and Allocating Costs to Consumers

Emerald Dudzinski, Virginia Tech, Blacksburg, VA, United States

This work optimizes supply chain decisions—inventory management, delivery scheduling, and vehicle routing—to enhance reliable access to safe drinking water for underserved communities. Additionally, we examine affordability and cost allocation to support sustainable operations. Partnering with a nonprofit, we leverage community-representative data and expert validation to ensure practical, scalable solutions.

3 - Modeling Forced Criminality and Intervention in Sex Trafficking Using Multi-Follower Bilevel Optimization

Xiaowei Guo, Clemson University, CLEMSON, SC, United States, Thomas Sharkey

We propose bilevel programming models to analyze sex trafficking networks involving multiple forced illegal activities, with traffickers being the upper-level and victims being the lower-level. The trafficker assigns quotas for victims while aiming to minimize their total risk, whereas victims aim to fulfill assigned quotas with minimal harm. Given that the lower-level victim problems may have multiple optimal solutions, we examine both optimistic and pessimistic bilevel formulations. We propose a novel approach to reformulate the bilevel problem into a single-level problem for both variants of the problem. The pessimistic bilevel problem reformulated by the proposed

method demonstrates a comparable difficulty to the optimistic variant. Experimental results show that our proposed reformulation method better performance compared to the traditional dual-based reformulation approach. We provide insights into how to effectively disrupt these networks.

MA11

Building A Level 3 A313

Empirical Healthcare: Addressing Disparities and Labor Shortfalls in Health Systems

Job Market Showcase

Job Market Showcase

Chair: Maria Ibanez, Kellogg School of Management at Northwestern University, Evanston, IL, United States

1 - Bridging the Nursing Gap: How Virtual Nurse Adoption Mitigates Workforce Shortages and Improves Inpatient Care

Blair Liu, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Yuqian Xu, Bradley Staats

The growing shortage of registered nurses (RNs) has led hospitals to explore innovative solutions. One such intervention is the Virtual Nurse (VN) model, which enables RNs to perform non-clinical inpatient tasks—such as admissions and discharges—via synchronous video communication. Using encounter-level data from a large healthcare system and a difference-in-differences design, we analyzed nearly 28,000 inpatient cases. VN adoption reduced length of stay (−7.18%) and 30-/60-day readmissions (−2.00%, −1.95%). Mechanism analysis shows improvements stem from shorter clinical service durations, especially when VN support occurs at admission. Effects are stronger under moderate congestion, with experienced VNs, and among high-need patients.

2 - Disparity in Maternal Care: the Role of Operational Factors in Eliminating the Gap

Max Yakovlev, Vanderbilt University, Nashville, TN, United States, Maria Ibanez

Annually, approximately 60,000 U.S. women experience severe maternal morbidity (SMM)—unexpected labor or delivery outcomes with serious short- or long-term health impacts. Many of these outcomes are preventable with appropriate care choices and applications. However, demographic and socioeconomic disparities in obstetric care and outcomes persist, and Black mothers consistently experience higher odds of labor complications. In this paper, we examine how hospital- and physician-level operational factors influence the racial disparity for Black mothers relative to other mothers in the likelihood of (1) riskier delivery procedures (e.g., first-time unscheduled Cesarean Section) and (2) resulting 90-day readmission rates following these procedures. Using cross-sectional data from all labor and delivery admissions across six U.S. states, we uncover key operational factors moderating these disparities: stronger hospital- and physician-level C-section specialization reduces the gap, as does greater exposure to treating Black patients. For instance, the odds of C-section for a Black mother are 11.1% lower when receiving care at a hospital in the top quartile for exposure to Black patients, compared to hospitals outside the top quartile. Notably, hospital ownership also impacts disparities: delivering a baby at a non-profit hospital is associated with lower odds of receiving a first-time unscheduled C-Section for a Black mother, whereas receiving treatment at a government-owned facility corresponds to higher odds of experiencing riskier procedures and 90-day readmissions. These findings highlight the importance of more efficient patient-provider matching, more detailed quality information disclosure, and stronger regulatory oversight. Such measures are crucial to promoting unbiased decision-making and advancing equitable health outcomes.

3 - Training Dilemma: Does More Certified Nurse Assistant Training Lead to Better Nursing Home Outcomes?

Catherine Wang, University of Wisconsin-Madison, Madison, WI, United States, Bob Batt, Hessam Bavafa

Certified Nurse Assistants (CNAs) are essential to the care of residents in nursing homes, yet concerns remain about whether current training requirements are sufficient. This study examines the effect of CNA training hours on nursing home quality and staffing. While federal regulations mandate a minimum of 75 training hours—including both clinical and classroom components—several states have enacted higher standards. The causal effect of these increased training hours on nursing home outcomes, however, remains unclear. Leveraging variation in state training requirements from 2009 to 2017, we use a difference-in-differences approach to estimate the effect of training hours on quality and staffing outcomes. Contrary to policy goals, increased training hours are associated with declines in quality, including an 18.6% rise in deficiency citations and adverse resident outcomes, such as a 3.3% increase in antidepressant medications and a 35.4% decline in reported skin rash cases. We also find shifts in staffing composition, with full-time CNA hours falling by 3.9% and part-time hours rising by 8.6%. These results underscore the complexity of designing effective quality improvement policies and carry implications for addressing persistent quality and workforce challenges in long-term care.

MA12

Building A Level 3 A314

Human Capital, Decision-Making, and Organizational Performance

Job Market Showcase

Job Market Showcase

Chair: Fanny Chen, Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands

1 - Health Insurance and Labor Productivity: Evidence from US Manufacturing

Yasaman Asayesh, Carlson School of Management, University of Minnesota, Minneapolis, MN, United States, Anant Mishra

Although health insurance is a critical issue at the business-societal interface globally, its impact on labor productivity remains unclear and contested among U.S. practitioners and policymakers. Recent studies suggest that government-sponsored health insurance improves healthcare access for a significant portion of the workforce and may enhance labor productivity. However, the overall impact remains uncertain, as such policies could also reduce work incentives. In this study, we examine how health insurance affects labor productivity in the U.S. manufacturing sector, where nearly one-third of workers rely on public safety net programs. We leverage Medicaid expansion as a quasi-experiment and conduct a difference-in-differences analysis using manufacturing productivity data from 2,554 counties across 49 states between 2010 and 2019. We find a notable positive average effect (3.2% to 4.1%) of Medicaid expansion on labor productivity. These gains are driven by increased manufacturing output rather than reductions in labor input. We further explore the mechanisms behind this effect and find that the productivity gains are not unconditional. Rather, the impact of Medicaid expansion depends critically on local health system capacity. Specifically, we focus on two dimensions: physical infrastructure (e.g., hospital beds per capita) and human resources (e.g., doctors and nurses per capita). Both are necessary conditions for realizing productivity gains. In counties with insufficient health system capacity, we find no significant effects. By identifying the conditions under which health insurance improves labor productivity, our findings offer actionable insights for policymakers seeking to align healthcare access with economic development.

2 - Quality Performance Feedback: the Role of Sequencing Feedback and Female Ownership on Service Outcomes

Peter Thurston Carrera, Fisher College of Business, Ohio State University, Columbus, OH, United States, Aravind Chandrasekaran, Kenneth Boyer

Research shows that performance feedback allows organizations to learn better. This study explores the effects of performance feedback content, sequence, and ownership diversity on learning. We use data from the hospitality industry to investigate these relationships and uncover that sequence and ownership diversity are key to improving service experience outcomes.

3 - Undue Managerial Influence in Selecting Suppliers for Organizational or Supervisory Benefit

Fanny Chen, Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands, Finn Wynstra, Jan van Dalen

While procurement norms explicate that supplier selection decisions should be free from undue influences, buyers often let these decisions be influenced by senior managers. Such undue influences intend to create benefits for the buying organization or manager. When listening to such influence, buyers engage in so-called unethical pro-social behaviors; actions that intend to benefit the organization or its members, but which violate norms and standards. Research has distinguished unethical pro-

organizational behavior and unethical pro-supervisor behavior but has not studied whether individuals behave differently depending on the intended beneficiary. Even though unethical pro-supervisor behaviors may carry a greater risk of sub-optimal results for the organization, unethical pro-organizational behaviors can have negative consequences in the long run too, particularly since affected suppliers may not care about intentions. We investigate whether undue managerial influence aimed at benefiting the organization has different effects on subordinates' supplier selection decisions compared to undue managerial influence aimed at benefiting the supervisor. We conduct two studies, combining between-subject manipulations with discrete choice experiments. The results show that unethical pro-organizational behavior significantly affects supplier selection decisions while unethical pro-supervisor behavior has limited significant effects on buyers' supplier selection decisions. Furthermore, limited significant differences exist between unethical pro-organizational and unethical pro-supervisor behaviors in supplier selection. This suggests that despite earlier arguments that these types of unethical behaviors are fundamentally different, buyers are nearly equally prone to unethical pro-organizational behavior and unethical pro-supervisor behavior. The results also show that moral identity does not curb these behaviors in supplier selection.

MA13

Building A Level 3 A315

Optimization Under Uncertainty II

Job Market Showcase

Job Market Showcase

Chair: Zihe Zhou, Purdue university, West Lafayette, IN, United States

1 - Stochastic Optimization with Multimodal Uncertainty and Decision Dependence

Shibshankar Dey, IEMS, Northwestern University, Evanston, IL, United States, Sanjay Mehrotra

Over the past decade, research on decision-making via optimization under distributional ambiguity has seen significant theoretical and computational advances. Recently, this work has expanded to learning-based decision problems that are grounded in optimization techniques while also informed by probabilistic generative modeling advances. Mixture models effectively represent unknown probability distributions, capturing multimodal uncertainties common in data from diverse sources. Motivated by these advances, this talk is organized into two parts.

The first part addresses optimization problems with probabilistic constraints, relevant to highly reliable decision-making under uncertainty. We focus on cases where the underlying uncertainty is modeled by Gaussian mixtures and show how these probabilistic constraints can be reformulated into tractable counterparts with and without distributional ambiguity using Wasserstein-based ambiguity sets.

In the second part, we consider the framework to incorporate decision-dependent forms of moment-based ambiguity sets and reformulate the resulting intractable models into workable forms. We show that state-of-the-art solvers highly underperform for such hard problems, whereas our adaptive approach delivers superior computational efficiency. Thus, this work contributes to both modeling techniques and algorithms for discrete decision-making under decision-dependent uncertainty.

2 - Data-Driven Optimization for Fast and Reliable Decision-Making Under Uncertainty

Irina Wang, Princeton University, Princeton, NJ, United States, Bartolomeo Stellato

Robust optimization and distributionally robust optimization have become widely adopted tools for building solutions that remain effective under uncertainty, but some challenges have always limited their practical functionality: high computational effort in terms of both memory and runtime, conservative models of the uncertainty leading to suboptimal solutions, and complex formulations that require advance knowledge of convex analysis. In this talk, I will address these challenges by discussing new data-driven frameworks for decision-making under uncertainty. In particular, I will focus on reducing data dimensionality to achieve computational speedups, and on learning contextual decision-focused uncertainty sets to reduce model conservatism.

3 - Drift Optimization of Regulated Stochastic Models Using Sample Average Approximation

Zihe Zhou, Purdue university, West Lafayette, IN, United States, Harsha Honnappa, Raghu Pasupathy

This work considers sample average approximation (SAA) of a general class of stochastic optimization problems over a function space constraint set and driven by "regulated" Gaussian processes. We establish statistical consistency by proving equiconvergence of the SAA estimator via a sophisticated sample complexity result. Next, recognizing that implementation over such infinite-dimensional spaces is possible only if numerical optimization is performed over a finite-dimensional subspace of the constraint set, and if sample paths of the driving process can be generated over a finite grid, we identify the decay rate of the SAA estimator's expected optimality gap as a function of the optimization error, Monte Carlo sampling error, path generation approximation error, and subspace projection error.

4 - Leveraging Graph Structure to Optimize Learning Algorithms

Mansi Sood, MIT, Cambridge, MA, United States

As our world becomes increasingly interconnected, the informational landscape that drives decision-making is marked by ever-expanding scale and complex interdependencies. Motivated by computational constraints in emerging computing paradigms, we study how better leveraging structure can push the frontiers of large-scale inference. Through the lens of graphical models, we develop computationally efficient alternatives to canonical subroutines that underlie inference in modern machine learning and optimization infrastructure.

We present our contributions along two key directions. First, we optimize graph algorithms for learning securely from distributed data sources, addressing a key challenge in decentralized settings—namely, identifying simple probabilistic rules for organizing nodes to balance sparsity with reliable connectivity. Our results resolve several open problems related to the exact analysis of connectivity properties in a class of random graph models known as random k-out graphs, widely appearing as heuristics for network design in settings with limited trust. Second, we develop computationally efficient alternatives to parameter estimation in graphical models that retain the statistical guarantees of classical methods like maximum likelihood estimation while significantly cutting computational costs. Our work sheds new light on how the interplay between graph structure and performance can be leveraged to push the frontiers of efficient and provably reliable algorithms.

MA14

Building A Level 3 A316

Diversity, Equity, Inclusion, & Doing Good

Contributed Session

Contributed

Chair: Sixiong You, Eli Lilly and Company

1 - Governing Higher Education Accessibility Knowledge Commons

Kyrie Zhixuan Zhou, University of Texas at San Antonio, San Antonio, TX, United States, Madelyn Sanfilippo

Lags in enforcement of Title II and market incentives to design for the majority and maintain default settings lead to significant challenges around accessibility (e.g., inaccessibility of PDF files, proprietary digital textbook platforms that are incompatible with screen readers) for college students with disabilities. In this mixed-methods case study, we combine auto-ethnography, a structured diary study among the authors, and computational content analysis of university policies to understand accessibility of higher education through the lens of Governing Knowledge Commons (GKC). GKC is a descriptive conceptual framework to support the analysis of community

governance of shared knowledge resources. The GKC framework allows us to unpack how shared knowledge and institutional structures are collectively managed to increase or decrease accessibility of higher education. Key action arenas identified in this study include decisions on procurement of accessible technologies (e.g., learning management systems), ineffective accommodations for students, underfunded disability services, overreliance on retrofitting rather than implementing accessibility at the onset, and overpromising of universal design. By identifying common failures in ensuring accessibility, we propose policy and design recommendations to rectify these issues, including: (1) accessibility training for faculty and staff members, (2) ensuring software accessibility in the procurement and operation processes, and (3) applying universal design principles consistently and contextually.

2 - DEI Commitment and Firms Operational Efficiency: Deep Learning Evidence from 300 million Job Posts

Shucheng MIAO, The Hong Kong Polytechnic University, Hong Kong SAR, Hong Kong, Xuemin Luo, Andy Yeung

Problem Definition: Diversity, Equity, and Inclusion (DEI) has become a strategic priority for organizations, yet its operational implications remain underexplored. This study addresses three questions: (1) Do job posts exhibit variations in DEI commitments, and if so, what characteristics lead to stronger DEI commitment? (2) Does DEI commitment relate to firm-level operational efficiency, and under what conditions will such effects be more pronounced? (3) Are these relationships robust to external societal events or regulatory shocks? **Methodology/Results:** Leveraging advanced deep learning BERT and RoBERTa algorithms, we construct a novel DEI commitment measure from over 300 million U.S. job posts, matched with firm-level financials for public firms. We find: (1) DEI language varies systematically across jobs, sectors, and years; (2) moderate DEI commitment enhances OE, especially in large, labor-intensive, and mature firms, but (3) an inverted U-shaped pattern emerges—excessive signaling erodes returns. Exogenous shocks of the BLM movement and Women on Boards Laws (SB-826) in California provide more causal and robust evidence for the inverted U-shaped relationship between DEI commitment and operational efficiency. **Managerial Implications:** DEI signaling can function as a dynamic capability when tuned to internal resources and external context. Yet over-commitment undermines cohesion and operational focus. Our DEI measure offers managers a data-driven deep learning tool to calibrate inclusion efforts: strategic alignment, not symbolic escalation, is key to sustaining operational efficiency performance.

3 - Performance and Misalignment: a Within-Subjects Study of New Role Integration in Hybrid Purpose-Oriented Networks

Xinmeng Huang, Northwestern University, Evanston, IL, United States

Facing more complicated societal, technical and operational challenges, organizations' integration of new external roles into existing teams to facilitate progress, communication, and innovation is becoming a standard and recurring practice. Meanwhile, new network structures, such as Multi-Team Systems, with cross-functional and cross-sector characteristics, are becoming increasingly prevalent. As AI has evolved in its societal role and taken on increasingly prominent responsibilities in high-stakes collaborative contexts, its integration into task-driven organizational settings is inevitable. As more teams in organizations welcome functional colleagues such as AI into their team, however, the impact on the existing teams' dynamics, objective alignments, and performance remains unanswered. Using within-subject designs, we collected data points from students at Northwestern University in a classroom study. Teams made an initial decision without AI, followed by decisions with the integration of AI teammates in different cross-functional network structures. Answers to cognition-based trust questions—inquiries built on rational evaluation of reliability or competence—were collected at the end of each stage. Our data-driven findings show that the integration of AI reduces cognition-based trust, alignments, and performance in existing teams. In contrast to traditional human teams, cognition trust using individual as trust referent is a stronger indicator of team performance than trust using team as trust referent in teams where AI was introduced. These findings lay the groundwork for more in-depth exploration of how integration of new roles, such as AI, influence team dynamics, alignments, and performance.

4 - Visualizing Cyber-Physical Attacks Through Physics-Informed Modeling

Madison Evans, Auburn University, Auburn, AL, United States, Greg Purdy

While connectivity has transformed manufacturing operations, it has outpaced the ability to adequately secure them. The interconnected and highly digitized manufacturing floor introduces unique cyber-physical vulnerabilities that could lead to serious physical consequences. Due to the risk of equipment damage, real-world data on cyber-physical attacks remains limited. To address this challenge, we are developing a Cyber-physical Manufacturing Range (CPMR) — a controlled environment, consisting of a dedicated machine and its digital counterpart. The physical machine informed the creation and movement of the digital replica, which was developed using the Unity real-time platform. However, repeating this process for attack scenarios is limited by the physical risks associated with executing live attacks. To generate realistic attack visualizations across multiple levels of fidelity, we must combine a variety of techniques. In this presentation, we will show how we are using these techniques to create varying types of physics-informed visualizations. We will discuss the shortcomings of those techniques and outline planned improvements to increase realism, scalability, and accuracy.

5 - Efficient Engagement: Opportunities for Mobile Methadone Maintenance in Geographically Underserved Areas

Anthony Bonifonte, Denison University, Granville, OH, United States, Erin Garcia

In 2021 the Drug Enforcement Agency changed regulations to allow easier registration and operation of mobile methadone clinics for treating opioid use disorder in rural and underserved areas. How these routes might best be operated and how many clients might be served by mobile units are open questions. This work identifies candidate dispensing locations from mobile units in areas that do not have geographic access to a fixed clinic. Optimal routes for mobile methadone are identified by solving a mixed integer program that maximizes the number of clients without access to care that can be served by a mobile route. Ohio is used as a case study state to explore the efficiency of routes leaving from each methadone clinic in the state. The approach is generalized nationwide, and multi-vehicle routes are analyzed. Between 48 and 68 clients can be served by a single vehicle in a day (mean of 60) under default parameter assumptions across Ohio. Similarly, nationwide routes serve a mean of 47.5 clients per day. The second and third vehicle routes from a clinic provide access to an additional 95% and 93% of the clients of single vehicle routes, respectively. The number of servable clients on any route greatly differs based on geography. This work demonstrates the potential value of mobile methadone maintenance in reaching populations without geographic access to a fixed clinic. Additional policy and funding changes are needed to incentivize the operation of mobile methadone units for serving rural and unmet demand.

6 - Participatory Objective Design via Preference Elicitation

Ali Shirali, University of California, Berkeley, Berkeley, CA, United States

In standard resource allocation, designers fix the objective function, assuming agents only care about their own allocations. This neglects agents' potential preferences over the objective itself—such as concerns for inequality or altruism—which cannot be captured by standard utility reports. We propose a participatory framework where agents express preferences over possible objectives, and the designer solves a modified problem that combines a benchmark objective with these preferences. We instantiate this in a fair division setting, where agents divide divisible goods and some are inequality-averse. Using the Fehr-Schmidt model, we show that this participatory approach can yield significant welfare gains, with bounded losses in the standard objective. Our theoretical results highlight tradeoffs and suggest that incorporating agents' preferences over objectives can outperform standard approaches in natural settings.

7 - Physics-Informed Predictions of Backorders in a Supply Chain

Sixiong You, Eli Lilly and Company, Indianapolis, IN, United States, Souptik Majumder, David Buchta

Accurate forecasting of back orders is critical for effective supply chain management and production planning. Accuracy is compromised by uncertainty, time delays, hidden variables, and nonlinear effects—common features in real-world supply chain systems. To enhance accuracy, we combine partial historical observations (e.g., supply, consumption, backorders) with physical constraints (conservation of mass for inventory) in a nonlinear dynamical model for backorders. This constraint guarantees physical law and reduces noise from data. The primary nonlinear model considered for demonstration is the Sparse Identification of Nonlinear Dynamical Systems (SINDy). As a data-driven approach that derives explicit ordinary differential equations (ODEs) from historical observations, SINDy provides transparent and parsimonious representations of system dynamics, enabling not only accurate forecasting of backorders but also revealing the underlying mechanisms — such as demand fluctuations and replenishment lags — that drive their evolution over time. This ODE-based formulation with time-delay embeddings is well-suited to capture the temporal dependencies and nonlinear dynamics commonly observed in supply chain systems. We evaluate our approach using a synthetic dataset, emulating a multi-echelon supply chain system, and benchmark it with classical ML models, including Linear Regression, Random Forest (RF), and XGBoost as well as domain agnostic foundational

models such as Chronos (Ansari et al. 2024). Our findings suggest that incorporating physical constraints into data-driven modeling approaches improves performance and contributes to more interpretable representations of supply chain dynamics

MA15

Building A Level 4 A401

Air Transportation Distinguished Speaker

Invited Session

Air Transportation Section

Chair: Max Li, University of Michigan, ANN ARBOR, United States

1 - Operational Approaches to Improve the Sustainability of Commercial Aviation

R. John Hansman, Massachusetts Institute of Technology, Cambridge, MA, United States

Operation approaches for reducing fuel burn and persistent contrail formation are the most effective near term response to the worldwide focus on improving the sustainability of air transportation. This presentation will first discuss potential approaches and then a worldwide comparison of 1 year of ADS-B surveillance data with fuel optimal trajectories to identify systemic areas of inefficiency. The analysis shows different mechanisms of inefficiency for horizontal, vertical and speed inefficiencies and variations between world regions. Potential approaches for improving the flexibility of air traffic control will also be discussed.

MA16

Building A Level 4 A402

Edelman Finalist Reprise

Invited Session

The Practice Section of Informs

Co-Chair: Irv Lustig, Princeton Consultants, Coral Gables, FL, 33146, United States

1 - Dynamic Route Optimization and Automation of Industrial Routes at Wm

Stuart Greene, WM, Houston, TX, United States, Hema Pillutla, Marcel Dalby, Seongbae Kim, Heesu Hwang, Sergiy Savchenko, Chao Zhou

WM, the leading provider of environmental solutions, faced significant challenges in optimizing its industrial waste collection routes following the pandemic. Traditionally, these routes were planned manually, a process that was time-consuming, labor-intensive, and suboptimal with multiple factors that need to be considered. WM has embarked on a journey to develop and implement a dynamic route optimization program tailored to the unique requirements of its industrial waste collection operations. The industrial waste collection sector presents a complex routing problem due to the dynamic nature of the customer service demand, various factors such as customers specific service requirements, container types, sizes, the disposal of waste materials among others. The dynamic route optimization system leverages data analytics, forecasting future service demand for effective capacity planning, advanced algorithms using metaheuristics to automate generation of routes that improve efficiency while ensuring safety and fulfilling customer service commitments. By effectively combining advanced analytics, optimization, and technology led automation for the industrial line of business, WM realized best-ever efficiency gains and operating margins and set the foundations for driving increased value in the future.

2 - Regionalize and Scale: Amazon Network Design for Faster and Cheaper Delivery

Cristiana Lara, Amazon, Bellevue, WA, United States, Ling Zhang, Shahbaaz Mubeen

By the end of 2021, Amazon in the US was selling over a billion items per year, through a network of hundreds of fulfillment centers, each carrying millions of unique items, spread throughout the continental US. These items were delivered through a single national transportation network consisting of internal resources and external delivery partners, leveraging a variety of facilities (fulfillment centers, sortation centers, and delivery stations), that was multi-modal and densely inter-connected. Amazon's proposition of offering Earth's largest selection, fast, with low cost-to-serve, faced challenges as a result of this unprecedented network growth. The need for speed requires systems that make real-time decisions on where to fulfill orders from based on the state of the network and inventory availability, which in turn results in high variability in the set of fulfillment centers (FCs) serving any destination area (e.g. city). That high variability translates to increased costs and reduced speeds.

In 2021, Amazon senior leadership identified this problem and assembled a team, the majority of whom were Operations Research (OR) professionals, to develop a solution. That solution is Regionalization, where we partitioned the US into 8 interconnected but largely self-sufficient regions. The development and launch of Regionalization were the result of 1.5 years of sustained OR studies examining each aspect of Regionalization, leveraging a multitude of OR methodologies such as mixed-integer programming, queueing theory, discrete-event simulation, inventory theory, machine learning, and statistical analysis. Regionalization was implemented in early 2023, and has since resulted in significant cost and speed improvements.

MA17

Building A Level 4 A403

Interpretable AI for Critical Decision-Making

Invited Session

Practice Track

Chair: Lili Zhang, Hewlett Packard Enterprise, Alpharetta, GA, United States

1 - Practices of Improving Accurate Interpretations of Graphrag with Structured and Unstructured Data

Lili Zhang, Hewlett Packard Enterprise, Alpharetta, GA, United States

Graph-based Retrieval-Augmented Generation (GraphRAG) is emerging as a cutting-edge technology that combines the structured relationship modeling capabilities of Knowledge Graphs with the contextual reasoning strengths of Large Language Models (LLMs). This results in a dynamic and efficient system for managing and generating information. However, like other GenAI or RAG applications, GraphRAG can face challenges related to accurate interpretations. In this talk, we will share practices for improving the accurate interpretations of GraphRAG by optimizing knowledge graphs with both structured and unstructured data. These practices are based on experimental comparison results across multiple use cases, such as customer engagements, user activities, and product diagnosis.

2 - Transparency and Interpretability in Supply Chain Inventory Optimization

Fan Wu, Intel, Atlanta, GA, United States, David Bayba

The emergence and swift advancement of AI and large language models (LLMs) significantly enhance the interpretability of supply chain recommendations derived from complex machine learning models. In the realm of inventory forecasting and management, we employed modern machine learning and time series methods to predict consumption over the next nine months and calculate key inventory metrics such as safety stock and reorder points based on these forecasts. Prior to the integration of LLMs, analysts spent considerable time investigating and communicating with stakeholders to ensure recommendations aligned with business practices and to explain the rationale behind forecasts generated from historical consumption patterns. With LLMs integrated into the system, a bot-like application can directly address stakeholders' inquiries by querying the knowledge base—such as databases containing historical consumption data and related features influencing forecasts—and making informed assumptions. Future work involves enhancing the model application to proactively explore and summarize use cases that stakeholders might categorize as anomalies but are currently unaware of. This approach aims to build a more robust inventory system and effectively monitor its overall health.

MA18

Building A Level 4 A404

Computational Foundations of Learning and Optimization in Multi-Agent Systems

Invited Session

Military and Security

Chair: Tao Li, City University of Hong Kong, Yeung Kin Man Academic Building, Hong Kong, N/A

Co-Chair: Quanyan Zhu, New York University, New York, NY, United States

1 - A Characterization of Reny's Weak Sequentially Rational Equilibrium Through Epsilon-Perfect Gamma-Weak Sequentially Rational Equilibrium and Its Selection

Chuangyin Dang, City University of Hong Kong, Kowloon, Hong Kong, Yiyin Cao

A weakening of sequential rationality of sequential equilibrium yields Reny's (1992) weak sequentially rational equilibrium (WSRE) in extensive-form games, which requires Kreps and Wilson's (1982) consistent assessment to satisfy global rationality of nonconvex functions at every information set reachable by a player's own strategy. The consistent assessment demands a convergent sequence of totally mixed behavioral strategy profiles and associated Bayesian beliefs. Nonetheless, due to the nonconvexity, the existence of WSRE had to invoke the existence of normal-form perfect equilibrium, which is sufficient but not necessary. Moreover, Reny's definition of WSRE contains insufficient information on constructing the convergent sequence. To address these two challenging issues, this paper develops a characterization of WSRE through ϵ -perfect γ -WSRE with local sequential rationality, which is achieved with the introduction of an extra behavioral strategy profile. For any given $\gamma > 0$, we generate a perfect γ -WSRE as a limit point of a sequence of ϵ_k -perfect γ -WSRE with $\epsilon_k \rightarrow 0$. A WSRE is then acquired from a limit point of a sequence of perfect γ_q -WSRE with $\gamma_q \rightarrow 0$. This definition allows one to analytically find all the WSREs for small extensive-form games and directly prove the existence of WSRE. An application of the characterization yields a polynomial system to serve as a necessary and sufficient condition for determining whether a totally mixed assessment is an ϵ -perfect γ -WSRE.

2 - Data-Driven Leader-Follower Game

Bo Zeng, Pitt, Pittsburgh, PA, United States, Zhengsong Lu

In this talk, we present a study on a data-driven leader-follower game where the utility function of the follower is known up to an ambiguity set. By employing the recent methodological development on distributionally robust optimization (DRO), this complex model can be solved efficiently, leading to a reliable solution for the leader, who might be a policy maker for a large number of followers.

3 - Distributed Dynamics and Stable Outcomes in Coalitional Games and Matching

Jeff Shamma, University of Illinois Urbana-Champaign, Urbana, IL, United States, Aya Hamed

Cooperative games model strategic settings in which agents coordinate to form partnerships and share payoffs to achieve mutually beneficial outcomes. These settings range from two-sided matchings to larger groups forming coalitions. A stable outcome is where no subgroup of agents has an incentive to break away to form a separate partnership. Stable outcomes, when they exist, can be computed in a centralized manner, where a single entity has full information about the underlying setup and full authority to assign partnerships and payoffs. Here, we present distributed dynamics where agents both form and break partnerships according to evolving internal aspiration levels that reflect self-interest. Our distributed dynamics require simple computations, limited memory, and most importantly, minimal knowledge of the environment. We establish that these dynamics converge to stable outcomes analogous to the core and present a variety of computational experiments.

4 - Conjectural Online Learning Dynamics Under Model Misspecification

Tao Li, City University of Hong Kong, Hong Kong, Hong Kong

As modern cyber-physical systems grow in scale and intricacy, their principled design, defense, and management depend on appropriate mathematical modeling, based on which computationally efficient operation strategies are developed. The fundamental challenge of this "modeling-then-optimize" paradigm arises from the entanglement between model **misspecification**, which is due to the epistemic uncertainty of the underlying system, and model **approximation**, which is intentionally introduced to trade model fidelity for computation efficiency. This work investigates the entanglement between misspecification and approximation in the context of partially observed Markov decision processes (Markov Games), where a decision-making agent holds a conjectured model of the system environment, which is calibrated using information feedback induced by decisions devised under the approximated conjecture model. The central question we aim to answer is whether the approximation affects model misspecification calibration. We consider Bayesian learning as the canonical approach for online model calibration and belief-quantization as the approximation scheme, and study the asymptotic behavior of the dynamical systems induced by such online learning, referred to as conjectural online learning. We study the self-fulfilling prophecy in conjectural learning dynamics by inspecting its attracting basin.

MA19

Building A Level 4 A405

Foundations of Generative Diffusion Models

Invited Session

Applied Probability Society

Chair: Renyuan Xu, Stanford University, Palo Alto, CA, United States

1 - Data-Driven Generative Simulation of Sdes Using Diffusion Models

Xunyu Zhou, Columbia University, New York, NY, United States, Xuefeng Gao, Jiale Zha

This paper introduces a new approach to generate sample paths of stochastic differential equations (SDEs) using diffusion models, a class of generative AI models commonly employed in image and video applications. Unlike traditional Monte Carlo methods, which require explicit specification of the drift and diffusion coefficients of the SDE, our method takes a model-free approach. Given a finite set of sample paths from an unknown SDE, we propose a data-driven framework that utilizes conditional diffusion models to generate new, synthetic paths of the same SDE. To demonstrate the effectiveness of our approach, we conduct comprehensive experiments on various SDEs and compare its performance with alternative benchmark methods including neural SDEs. Furthermore, we explore the potential of leveraging these synthetically generated sample paths to enhance the performance of reinforcement learning algorithms in continuous-time mean-variance portfolio selection, hinting promising applications of diffusion models in financial analysis and decision-making.

2 - Dimension-Free Convergence of Diffusion Models for Approximate Gaussian Mixtures

Yuting Wei, University of Pennsylvania, PHILADELPHIA, PA, United States

Diffusion models are distinguished by their exceptional generative performance, particularly in producing high-quality samples through iterative denoising. While current theory suggests that the number of denoising steps required for accurate sample generation should scale linearly with data dimension, this does not reflect the practical efficiency of widely used algorithms like Denoising Diffusion Probabilistic Models (DDPMs). In this talk, we investigate the effectiveness of diffusion models in sampling from complex high-dimensional distributions that can be well-approximated by Gaussian Mixture Models (GMMs). For these distributions, our main result shows that DDPM takes at most $O(1/\epsilon)$ iterations to attain an ϵ -accurate distribution in total variation (TV) distance, independent of both the ambient dimension d and the number of components K , up to logarithmic factors. Furthermore, this result remains robust to score estimation errors. These findings highlight the remarkable effectiveness of diffusion models in high-dimensional settings given the universal approximation capability of GMMs, and provide theoretical insights into their practical success.

3 - Generator Matching: Unified Theory for Generative Models

Minshuo Chen, Northwestern University, Evanston, IL, United States

Generative AI has achieved remarkable success in modeling high-dimensional, multi-modal data through techniques such as diffusion models and normalizing flows. However, the theoretical underpinnings of these diverse modeling paradigms remain fragmented and underdeveloped. We close the gap via *generator matching*, a unified framework that encompasses autoregressive modeling in LLMs, denoising score matching in diffusion models and velocity matching in flow-based generative models. By interpreting both as instances of matching the generator of a continuous-time Markov process to model the data, this framework generalizes and extends current approaches to modeling complex, multi-modal distributions. We establish statistical learning theories for generator matching, providing consistency guarantees and generalization bounds. Our theory not only bridges existing paradigms in generative modeling but also opens new avenues for principled design of stochastic processes underlying modern generative models.

4 - Computational Bottlenecks for Denoising Diffusions: An Example of Sparse and Low-Rank Distributions

Viet Vu, Stanford University, Stanford, CA, United States

Denoising diffusions sample from a probability distribution μ

in $\mathbb{R}^d$ by constructing a stochastic process $(\mathbf{x}_t)_{t \in [0,1]}$ in $\mathbb{R}^d$ such that $\mathbf{x}_0$ is easy to sample, but the distribution

of $\mathbf{x}_T$ at large T approximates μ . The drift

$\mathbf{m} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ of this diffusion process is learned by minimizing a

score-matching objective.

Is every probability distribution μ , for which sampling is tractable, also amenable to sampling via diffusions?

We provide evidence to the contrary by studying a probability distribution

μ for which sampling is easy, but the drift of the diffusion process is intractable ---under a popular conjecture on information-computation gaps in statistical estimation.

We show that there exist drifts that are superpolynomially close to the optimum value

(among polynomial time drifts) and yet yield samples with distribution that is very far from the target one.

MA20

Building A Level 4 A406

RAS Problem Solving Competition

Award Session

Railway Applications

Chair: Marty Schlenker, Schlenker LLC, Mechanicsburg, PA, 17055, United States

MA21

Building A Level 4 A407

From Streets to Systems: Cross-Domain Digital Twins in Urban Transportation

Invited Session

Transportation Science and Logistics (TSL)

Chair: Zilin Bian, New York University, Brooklyn, United States

Co-Chair: Tao Li, City University of Hong Kong, Yeung Kin Man Academic Building, Hong Kong, N/A

1 - Digital Twins for Transportation and Infrastructure: Accelerating Scenario Creation and Real-TIME Analysis

Ruimin Ke, Rensselaer Polytechnic Institute, Troy, NY, United States

This presentation unifies three recent advances from the speaker's group into a cohesive digital-twin pipeline suitable for campus-scale transportation studies. First, the ChatSUMO agent converts plain-language requests (e.g., "medium traffic within a one-mile radius of Albany") into complete SUMO networks and demand files in a short time, ten-to-fifteen times faster than manual scripting, without requiring simulation expertise. Second, an edge-computing module deploys detection and tracking methods on a Jetson Xavier NX edge computer to detect and track vehicles in real-time, map their 2-D image coordinates into 3-D space, and stream the results to a CARLA-based campus digital model, overcoming occlusion and lens-distortion challenges to maintain a live vehicle layer in the twin. Finally, a CARLA-SUMO co-simulation environment ingests those vision feeds and applies multi-agent Q-learning for adaptive signal timing, providing a test bed that marries realistic rendering with control algorithms. Together, these components demonstrate a practical, data-driven workflow in which analysts can describe a scenario in natural language, populate it automatically, tether it to real-time edge sensing, and evaluate optimization strategies. The talk will discuss methodology, validation results, and lessons in rapidly reconfigurable digital twins aimed at transportation decision support.

2 - Mobility Digital Twin with AI Agents and AI Networks

Xishun Liao, University of Central Florida, Florida, FL, United States

Mobility Digital Twins (MDT) are rapidly evolving into sophisticated tools for simulating and optimizing smart city systems. In this talk, a novel MDT framework built on the integration of the concepts of AI Agents and AI Networks will be introduced. Within this framework, AI Agents serve as intelligent digital replicas of individual entities, such as travelers, drivers, vehicles, capable of learning, decision-making, and generating behavioral patterns at varying resolutions to capture nuanced cross-domain interactions. These agents dynamically adjust to evolving conditions in the AI Networks and can simulate complex behaviors that are critical for realistic mobility modeling. Concurrently, AI Networks represent a multi-layered digital ecosystem that encompasses both dynamic components, such as traffic flows and sensor data, and static elements of the built environment, including roadway networks, point-of-interest, and infrastructures utilities. Learning, generative, and simulation techniques play a critical role in building, maintaining, and updating these interconnected layers. Together, the Mobility AI Agents and Networks enable (near) real-time and offline prediction, decision-making, planning, optimization, and simulation, supporting a wide range of smart mobility applications.

This talk will elaborate on key modules required to develop the multi-layer multi-resolution MDT and the technical challenges ensuring model scalability and adaptability and maintaining computational efficiency. Additionally, practical applications of the MDT framework in transportation system will be discussed. Through these insights, the talk will demonstrate how MDTs enhance the responsiveness, efficiency, and sustainability of urban transportation networks.

3 - Leveraging Affordable Cav Edge Devices to Assist Digital Twin Modeling

Hao Zhou, University of South Florida, Tampa, FL, United States, Shengming Yuan

The widespread adoption of digital twin (DT) technology in transportation is currently constrained by its reliance on expensive, dedicated roadside sensors and communication infrastructure. This research addresses this critical gap by leveraging the untapped potential of increasingly prevalent in-vehicle sensors such as ADAS-equipped cars and user dash-cams, to offload infrastructure costs and enrich DT models. Our work bridges on-vehicle Cooperative Driving Automation (CDA) edge capabilities with sophisticated DT modeling through a framework for comprehensive data acquisition from vehicle CAN bus and external cameras. A cornerstone of our approach is the utilization of low-cost communication strategies—specifically next-generation Wi-Fi (Wi-Fi 6/7) and existing cellular networks—which obviate the need for substantial new infrastructure investments, complemented by the resilient QUIC protocol and WebP image compression. We employ Generative AI for intelligent data post-processing tailored to DT modeling standards. Critically, to compensate for the potential scarcity of dedicated edge CDA devices, we demonstrate that vision transformers can reconstruct comprehensive road scenes for DT modeling from vehicle-derived visual data, mitigating concerns about the scarcity of CDA edge deployment. This integrated strategy, by transforming vehicles into primary data contributors empowered by AI and cost-effective communication, significantly lowers the entry barrier, enabling scalable and accessible DT deployment.

MA22

Building A Level 4 A408

Data-Driven Operations

Invited Session

Applied Probability Society

Chair: Yueyang Liu, Rice University, Houston, United States

Co-Chair: Chao Qin, Stanford University, Palo Alto, CA, United States

1 - On Information Design with Spillovers

Yifan Guo, Stanford University, Stanford, CA, United States, Ozan Candogan, Haifeng Xu

An information designer chooses which experiments to assign to agents in a directed network, where agents also observe spillovers from others' experiments. We study optimal information design under different spillover modes (local vs. global) and under lossy transmission. In the global case, when each agent follows at most one other, the optimal structure is tractable; but if some follow more than one, the problem becomes NP-hard, even with simple communication technologies. Local spillovers show a different pattern: the problem remains tractable if agents follow only few others and communication is simple, but any deviation leads to intractability. Finally, we apply our framework to supply chains, where a retailer reveals demand information to affect upstream investment. By exploiting payoff structure, we obtain tractable algorithms in richer settings and highlight how different spillover types shape the retailer's payoff.

2 - Variance Aware Thompson Sampling for Expediting A/B/N Testing

Shuangning Li, University of Chicago, Chicago, IL, United States

In many experimental settings, the goal is to compare a baseline (control arm) with multiple variations (treatment arms) and maximize the probability of identifying at least one treatment that outperforms the control under a fixed sample size. The standard approach involves A/B/n testing, where experimental units are randomly assigned to the control or one of the treatment groups, followed by statistical testing (e.g., t-tests) to assess significance.

We focus on scenarios where units arrive in batches and the inference procedure—batchwise t-tests followed by aggregation—is fixed, often due to managerial or regulatory constraints. While the inference method must remain unchanged, allocation probabilities across arms can be adaptively adjusted before each batch, framing this as an optimal experimental design problem.

To this end, we develop a variance aware Thompson Sampling method. Unlike standard Thompson Sampling, our approach prioritizes treatment arms with both high expected reward and low variance, and allocates more samples to the control arm to reduce variance in comparisons. Compared to uniform randomization, this method avoids wasting samples on unpromising arms and expedite the process of identifying superior treatments under the fixed inference framework.

3 - New Relative Value Iteration Algorithms for the Ergodic Risk-Sensitive Control of Markov Chains

Nicola Sassone, Department of Computational Applied Mathematics and Operations Research, Rice University, Houston, TX, United States, Sumith Anugu, Guodong Pang

In this paper, we propose new Jacobi-like and Gauss-Seidel-like relative value iteration (RVI) algorithms for the ergodic risk-sensitive control (ERSC) problem of a controlled discrete-time Markov chain (DTMC) on discrete state space and establish their convergence. With finite state space, whenever the DTMC is irreducible and recurrent under every stationary Markov control policy, we prove that the iterates of our algorithms converge geometrically. The proof requires establishing a local contraction property, and also a local Lipschitz continuity property of the fixed point of the risk-sensitive Bellman-like operator. To tackle the challenges associated with the multiplicative nature of the operator, we employ the well-known entropy variational formula, which re-casts the operator and its fixed point equation in an additive form, albeit with an additional optimization problem over a corresponding set of probability vectors. In the case of countable state space, we propose an implementing procedure by properly truncating the problem to have finite state space. Under either uniform stability or a near-monotonicity condition, we show that as the truncation size grows indefinitely, the iterate values of our algorithms converge to the value function and the optimal ERSC cost. Finally, we implement our algorithms on two examples to demonstrate the performances of our proposed RVI algorithms: maximizing the exit rate from a finite domain (on a graph) and a single-server queue of finite capacity in the case of finite state space, and also a single-server queue of infinite capacity with abandonment in the case of countable state space.

4 - Calibration in Heavy-Tailed Machine Learning: a Statistical Perspective

Wenhao Yang, Stanford University, Stanford, CA, United States, Jose Blanchet, Peter Glynn, Aleksandar Mijatovic

Heavy-tailed noise has gained significant attention in modern machine learning due to its impact on generalization performance in stochastic training algorithms, especially stochastic gradient descent (SGD). While empirical evidence suggests that heavy-tailed noise can enhance optimization dynamics and improve model generalization, it also poses fundamental challenges for theoretical analysis and statistical inference.

In this talk, I will present a statistical perspective on the role of heavy-tailed noise in SGD, focusing on two key aspects. First, I will discuss the asymptotic properties of SGD under heavy-tailed noise, highlighting key differences from classical light-tailed theory. I will then introduce a principled approach to statistical inference for SGD-based estimators, enabling uncertainty quantification despite the presence of heavy tails. These results provide new insights into the calibration of learning algorithms under non-standard conditions, bridging the gap between optimization and statistical inference. I will conclude by discussing the implications of these findings for practical machine learning applications, highlighting open questions and future research directions in the intersection of optimization, statistical inference, and robustness in heavy-tailed settings. This talk is based on joint works with Jose Blanchet, Peter Glynn and Aleksandar Mijatović.

5 - Managing Hidden Risk in Operations: Safety and Resilience

Feng Zhu, MIT, Cambridge, MA, United States, David Simchi-Levi

This talk is about understanding and managing hidden risk in various operational decision-making environments that bear a sequential manner.

Our first focus is managing tail risk in online learning, with applications in revenue management, clinical trials, and financial investment. We give a detailed characterization of regret distribution and demonstrates how managing regret tail risk leads to extra benefits on policy robustness. In Part I, we design policies that enjoy both optimal regret expectation and light-tailed regret distribution. In Part II, we study the optimal trade-off between expectation and tail risk for regret distribution.

Notably, we find that our policy design coincides with what was adopted in AlphaGo MCTS, providing high-level insights to the success of their engineered solution in complex decision-making environments.

Our second focus is managing disruption risk in supply chains. We introduce a new supply chain resilience framework that accommodates the impact of uncertain Time-To-Recover (TTR) on risk detection, response coordination, and system recovery across general supply chains. We propose a new Risk Exposure Index for pre-disruption risk detection, characterize the price of TTR uncertainty during the intra-disruption stage via a tight theoretical analysis, and include post-disruption system recovery by considering Time-To-Recover-Inventory (TTRI). An extensive numerical study supplements our analytical findings and provides new managerial insights.

MA23

Building A Level 4 A409

Life as a Ph.D. Student

Panel Session

Minority Issues Forum

Co-Chair: Meen-Sung Kim, Georgia Institute of Technology, Atlanta, United States

Co-Chair: Valentina Castaneda-Torres, Georgia Institute of Technology, 251 10th street NW, Atlanta, GA, 31308, United States

1 - Moderator Panelist

Meen-Sung Kim, Georgia Tech, Atlanta, GA, United States

The journey of pursuing a Ph.D. can be exciting, frustrating, rewarding, and exhausting, often all at once. In this panel, current Ph.D. students will share their personal stories: what led them to pursue a Ph.D., the challenges they've faced, and the lessons they've learned along the way. We'll talk about what it's really like to be a Ph.D. student—beyond the research papers and deadlines. Whether you're thinking about grad school or already in the middle of your program, we invite you to join us for an open, honest, and supportive conversation about navigating life as a Ph.D. student.

2 - Moderator Panelist

Valentina Castaneda-Torres, Georgia Institute of Technology, Atlanta, GA, United States

The journey of pursuing a Ph.D. can be exciting, frustrating, rewarding, and exhausting, often all at once. In this panel, current Ph.D. students will share their personal stories: what led them to pursue a Ph.D., the challenges they've faced, and the lessons they've learned along the way. We'll talk about what it's really like to be a Ph.D. student—beyond the research papers and deadlines. Whether you're thinking about grad school or already in the middle of your program, we invite you to join us for an open, honest, and supportive conversation about navigating life as a Ph.D. student.

3 - Panelist

Hasini Balasuriya, University of Louisville, Louisville, KY, United States

4 - Panelist

Pavithra Sripathanallur Murali, George Mason University, Fairfax, VA, United States

5 - Panelist

Fernando Acosta-Perez, University of Wisconsin-Madison, Madison, WI, United States

6 - Panelist

Alina Gorbunova, Georgia Institute of Technology, Atlanta, GA, United States

7 - Panelist

Pardis Bahmani, University of Missouri- St.Louis, Saint Louis, MO, United States

MA24

Building A Level 4 A410

TSL Dissertation Award and TS Journal Paper of the Year Awards

Award Session

Transportation Science and Logistics (TSL)

Chair: Hai Wang, Singapore Management University, Singapore, Singapore

1 - Intro to the TSL Dissertation Award

Hai Wang, Singapore Management University, Singapore, Singapore

This award is the oldest and most prestigious honor for doctoral dissertations in the transportation science and logistics area.

2 - Machine Learning for Sequential Decisions in Logistics

Fabian Akkerman, University of Twente, Enschede, Netherlands

This presentation will be given by a winner of a TSL award.

3 - Analytics for Better Urban Cycling-

Bo Lin, University of Toronto, Toronto, ON, Canada

This presentation will be given by a winner of a TSL award.

4 - Combinatorial Optimization-Enriched Machine Learning to Solve the Dynamic Vehicle Routing Problem with TIME Windows

Kai Jungel, Technical University of Munich, Munich, Germany, Axel Parmentier, Leo Baty, Maximilian Schiffer, Patrick Klein

This presentation will be given by a winner of a TSL award.

5 - Electric Vehicle Scheduling in Public Transit with Capacitated Charging Station

Rolf van Lieshout, Eindhoven University of Technology, Eindhoven, Netherlands, Marelou H de Vos, Twan Dollevoet

This paper considers the scheduling of electric vehicles in a public transit system. Our main innovation is that we take into account that charging stations have limited capacity, while also considering partial charging. To solve the problem, we expand a connection-based network in order to track the state of charge of vehicles and model recharging actions. We then formulate the electric vehicle scheduling problem as a path-based binary program, whose linear relaxation we solve using column generation. We find integer feasible solutions using two heuristics: price-and-branch and a diving heuristic, including acceleration strategies. We test the approach using data from the concession Gooi en Vechtstreek in the Netherlands, containing up to 816 trips. The diving heuristic outperforms the other heuristic and solves the entire concession within seven hours of computation time with an optimality gap of less than 3%.

MA25

Building A Level 4 A411

Assisting Delivery Via Drones or Other Aides

Invited Session

TSL: Freight Transportation

Chair: Mengting Chao, University of Maryland-College Park, College Park, MD, United States

1 - A Dynamic Replanning Strategy for Disruption-Aware Truck and Drone Deliveries**Mehmet Kurt, Auburn University, Auburn, AL, United States**

This study explores the potential effects of stochastic disruptions, such as adverse weather and unexpected events, on the well-known Flying Sidekick Traveling Salesman Problem (FSTSP). This work introduces the first FSTSP extension that incorporates dynamic recourse planning for disruption, addressing both total and rolling disruptions. We use a dynamic re-planning strategy, where we update the entire solution based on Bayesian updates to the expected start time and duration of the disruption each time a vehicle reaches a customer location, with additional re-planning when the disruption starts and ends. This approach accounts for the stochastic nature of both disruption start times and durations with assumptions of probability distributions of these and conditioning on the current time. We follow an iterative process: we generate the initial plan, determine the next decision epoch, move to the next closest node, roll the planning horizon forward, and replan for the entire remaining day from that point—continuing this cycle until the end of the day or until a disruption is encountered and a disruption replan is executed. The mathematical model is implemented to evaluate the mitigating effect of recourse actions, with computational experiments and a comparative analysis of rolling and total disruptions. A heuristic approach is developed to handle large-scale instances. Overall, this study highlights the importance of considering disruptions in truck-and-drone delivery planning. Through a dynamic update approach that incorporates stochastic disruption details, it demonstrates how real-time adjustments can improve the efficiency and resilience of last-mile delivery operations.

2 - Target-Driven Initial Tour Construction for the Traveling Salesman Problem with Drone**Jae Hyeok Lee, KAIST, Daejeon, Korea, Republic of, Minwoo Kim, Kiyoung Um, Jinkyoo Park, Changhyun Kwon**

This research challenges a fundamental assumption in the Traveling Salesman Problem with Drone (TSP-D) solution methodologies that optimal TSP tours provide the most effective foundation for high-quality TSP-D solutions. The TSP-D involves a truck and drone collaboratively serving customers while minimizing completion time. It represents a class of problems for coordinating heterogeneous autonomous agents in logistics. Current approaches typically generate a near-optimal TSP tour and then apply dynamic programming-based partitioning to determine drone operations. Through systematic investigation, we demonstrate that this approach often yields suboptimal results. Our comparative analysis reveals that tours with specific structural characteristics—particularly configurations with pronounced angular regions—consistently produce superior TSP-D solutions despite having higher TSP costs. These structures enable more efficient drone utilization by strategically positioning launch and recovery points. Building on this insight, we develop a reinforcement learning framework that directly optimizes initial tours for TSP-D performance rather than TSP optimality. The model formulates tour construction as a sequential decision process where rewards derive from the quality of resulting TSP-D solutions after exact partitioning. Our experiments demonstrate that RL-generated tours significantly outperform those derived from optimal TSP tours when evaluated as TSP-D solutions. This work reveals a methodological blind spot in sequential optimization approaches that extend beyond TSP-D to other multi-stage problems. By reconceptualizing the initial tour generation process to target downstream performance objectives directly, we achieve substantial improvements in solution quality without increasing computational complexity. We advance the state-of-the-art truck-drone routing optimization and establish a framework for target-driven construction heuristics in complex optimization problems.

3 - Two-Echelon Collaborative Delivery Through Truck-Drone Synergy**Linxuan Shi, The George Washington University, Washington, DC, United States**

Emerging drone technologies have driven interest in last-mile delivery, offering advantages such as flexibility, fast delivery, low cost, and a reduced carbon footprint. However, drone-based delivery services face constraints, including short delivery range and limited carrying capacity, which have led logistics carriers to explore hybrid truck-drone collaboration. This study evaluates the efficiency of a hybrid delivery paradigm in which a truck travels along major arterials while multiple drones sidekick within a designated strip area alongside the truck's trajectory to perform final package deliveries. We model this problem using continuous approximations and estimate two critical system performance metrics—truck traversal time and energy consumption—through a capacitated queuing model. Theoretical analysis and numerical results indicate that system performance is highly sensitive to the way that the truck and drones coordinate, particularly the drones' serviced strip width and the speed synergy between the two modes. Findings also suggest that truck-drone collaboration outperforms a truck-only delivery mode in both efficiency metrics, with performance improving as more drones are deployed.

4 - The Generalized Driver-Aide Problem: Enhancing The Role of The Aide in Urban Last-Mile Delivery**Mengting Chao, University of Maryland-College Park, College Park, MD, United States**

The dramatic growth in package delivery volume in recent years has presented significant challenges for last-mile delivery, especially in urban areas. We introduce the Generalized Driver-Aide Problem (GDAP), where an aide can assist the driver in two modes: (i) “Jumper” mode: The aide works with the driver to expedite the delivery at a stop, and (ii) “Helper” mode: The aide is dropped off by the truck to serve multiple stops independently while the driver serves other stops and picks up the aide later. This generalized helper mode allows the aide to move between multiple stops independently, a unique feature in dense urban settings. We must determine both the delivery route and the most effective way to use the aide to minimize the total delivery time. Our primary managerial question involves exploring the benefits of the aide's independent movement in urban last-mile delivery. We develop exact approaches to solve the GDAP. First, we provide a compact MIP formulation, which can be solved directly by off-the-shelf solvers. We then develop an IP formulation with exponentially many variables and constraints. Built on this formulation, we propose a sophisticated branch-cut-and-price approach with several algorithmic enhancements. We conduct a case study involving New York City, based on real-world data and the established literature. Comparing to the traditional truck-driver delivery without the aide, the GDAP leads to significant improvements in delivery efficiency, cost reduction, environmental sustainability, and mitigation of parking-related issues.

5 - Incentive-Compatible Mechanism for Integrated Bike Rebalancing and Local Store Deliveries**Nastaran Tork, University of Minnesota, Minneapolis, MN, United States, Alireza Khani**

This study proposes an incentive-compatible framework that enlists regular bike sharing users to perform either station rebalancing or last-mile deliveries within a single, daytime platform. Users disclose walking and biking tolerances, receive bundled task offers generated by a binary optimization model, and bid for their preferred task in a reverse second-price auction. Payoffs reflect a weighted task value that blends inventory impact with an early-delivery bonus. A two-hour simulation with 2017 Nice Ride Minnesota data confirms the framework's practicality. Assuming a 95% participation rate, users assigned to rebalancing tasks walked only 8% less than usual while biking 38% less. Every assigned delivery met its deadline, and the redistribution of bikes noticeably narrowed the gap between under- and over-stocked stations. Budget analysis shows that adjusting the weights of the task value factors strongly influences which tasks receive funding and, consequently, the system's overall benefit. These findings suggest that crowdsourced daytime rebalancing, paired with on-bike deliveries, can relieve fleet imbalances and help to cut vehicle miles with minimal user burden.

MA26

Building A Level 4 A412

Recent Advance in Theory and Applications of Facility Location

Invited Session

Location Analysis

Chair: Alan Murray, University of California at Santa Barbara, Santa Barbara, CA, United States

1 - Modified Sets of Closest Assignment Constraints for R-Hub Median Interdiction Problems**Saikat Mukherjee, Indian Institute of Technology Delhi, New Delhi, India, Prasanna Ramamoorthy**

Interdiction of hubs shows vulnerability within a hub network explaining how unanticipated occurrences or attacks can affect the network. Routing of flows can be possible through surviving hubs in post-interdiction situations. This study includes closest assignment constraints, which ensure allocation of non-hub nodes to the nearest hub nodes. In this study, hub median interdiction problem is modeled as bilevel optimization problem. These assignment constraints can be used to reformulate bilevel model into single level. We propose several alternate modified sets of closest assignment constraints. We analyze those constraints and present dominance relations among them. Experiment Results indicate that the effectiveness of constraints can create a tighter model which reduces computational complexity.

2 - A Bilevel Optimization Framework for Strategic Deployment of Survivor Geolocation Systems and Rescue Teams in Post-Disaster Operations

Jose Walteros, University at Buffalo, Buffalo, NY, United States, Angelo Joseph Soto Vergel, Diana Ramirez-Rios, Juan Borrero

The prompt collection of accurate information about potential survivor locations in the aftermath of a disaster is critical for the success of search and rescue operations. Traditional methods rely on teams combing through affected areas for signs of survivors. However, limitations on data availability, resources, and personnel can severely restrict the effectiveness of these efforts. Recent technological advancements have enabled the strategic deployment of remote sensors in affected areas, providing an effective tool for identifying signs of survivors before rescue missions are launched. The data collected by these sensors can significantly enhance the situational awareness of rescue planners, allowing them to guide rescue teams toward regions with potentially high survivor densities, thereby improving the likelihood of successful outcomes. This presentation introduces a bilevel stochastic optimization model with recourse that addresses the joint problem of placing sensor probes and deploying rescue teams. In the first level, the model selects candidate zones for deploying a set of geolocated probes equipped with various sensors (e.g., heat, sound, motion), subject to a predefined budget. These decisions rely solely on prior probabilities of the survivors' locations, inferred from pre-disaster indicators, hazard exposure risks, and socio-demographic markers. These factors serve as a general indicator of the potential presence of survivors and their level of endangerment. In the second level, the model determines the areas the teams will traverse, using the partial information collected from the deployed probes. The overarching objective of the model is to maximize the expected number of survivors that can be rescued.

3 - Dispersed and Compact Locational Siting

Alan Murray, University of California at Santa Barbara, Santa Barbara, CA, United States

The selection of areas for activity has long been of interest in location modeling. Such an area may represent a residential development, industrial processing plant, storage site, nature preserve, harvest activity and/or mitigation region, among others. Many important spatial features are often necessary or desired, including compactness and dispersion. This paper reviews approaches to accomplish these goals and introduces linear formulations to address associated needs.

MA27

Building B Level 2 B201

Recent Advances in Network Optimization

Invited Session

Computing Society

Chair: Haoming Shen, University of Arkansas, Fayetteville, AR, United States

Co-Chair: Robert Curry, University of Arkansas, Fayetteville, AR, United States

1 - Counter-Interdiction via Finite-Adaptability

Ningji Wei, Texas Tech University, Lubbock, TX, United States, Zolykha Rezaei

Interdiction games are Stackelberg games in which the attacker acts as the leader, selecting an optimal disruption plan, while the defender, as the follower, responds only after observing the attacker's actions. This sequential structure places the defender at a strategic disadvantage, as any fixed or predefined response can potentially be anticipated and neutralized by the attacker. This project shifts the focus to the defender's perspective by introducing the concept of a counter-interdiction game, where the defender is equipped with a contingent response plan that adapts after observing the attacker's decision. We develop several solution methods for this problem, supported by analytical performance guarantees. Beyond interdiction, the proposed methodologies contribute to a broader framework for robust finite-adaptability optimization, enabling robust defense strategies not only against adversarial attacks but also under general uncertainty.

2 - A Robust Bilevel Network Interdiction Problem with Applications in Human Trafficking Disruption

Daniel Lopes da Silva, Clemson University, Clemson, SC, United States, Thomas Sharkey, Yongjia Song

We introduce a robust bilevel network interdiction problem inspired by applications in human trafficking disruption. In this problem, the follower, who operates the network, solves a minimum cost flow problem. The leader, who interdicts the network, minimizes the total flow on the arcs from a special set in the follower's minimum cost flow solution. The problem includes data uncertainty as the leader does not know the follower's operational cost vector. The follower has complete knowledge about its own parameters and therefore makes decisions in a "wait-and-see" manner. We present a solution approach to the optimistic version of the problem and discuss interesting observations.

3 - Mixed-Integer Programming Models for Solving Semi-Cooperative Attack Graph Problems

Phanuel Allaissem Beremadji, University of Arkansas, Fayetteville, AR, United States

In this work, we focus on the Semi-Cooperative Attack Graph Problems (SCAGP), where we aim to optimize attack strategies in interdependent networks. In our problems, network information such as the number of nodes or states, the number of arcs or exploits, is known and shared among attackers. Additionally, attackers, with their parameters and utilizing their resources, collaborate to compromise the network. We proposed mixed-integer programming (MIP) models and a Lagrangian-based decomposition model to analyze and solve the problem. The decomposition method uses a relaxed MIP as a starting point and iteratively adds valid cuts to the relaxed MIP whenever a constraint is violated. We compare the computational results of the two models on various instances with different parameters, including the size of the networks, the number of attackers, the budget, and the cost of completing exploits. For future works, we may consider alternative objective functions, including minimizing the maximum lateness or weighted total tardiness.

The SCAGP model can be applied to various domains, including defense strategies for interdependent infrastructures, wireless sensor network security, and the spread of disinformation over social networks. By considering cooperative attack scenarios, our models contribute to a more comprehensive understanding of attack dynamics and optimization in real-world settings.

4 - Mixed Integer Programming Model for Robust Multi-Commodity Flow Scheduling Problems

Md Sazid Rahman, University of Arkansas, Fayetteville, AR, United States, Robert Curry

In this presentation, we study variant of a network flow maximization problem called the Multi-Commodity Maximum Value Dynamic Flow Problem (MC-MVDFF). In this problem, we maximize the value of the multi-commodity network flow schedule on a dynamic network, where an amount of residing minimum flow of a commodity acquires value on each node at the end of a time period. We assume flows traverse the network during each time period and reside at nodes between periods. We solve a mixed-integer programming (MIP) formulation for the problem considering uncertainty on the unit cost of flow and minimum amount required to acquire value for a commodity at a node. Moreover, we robustify the MC-MVDFF model (R-MC-MVDFF) with uncertainty and add two sets of constraints to the robust model for practicality: 1) we consider a set of permanent commodities for which if we acquire value once at a node for some time period, then we acquire value for that commodity and at that node for every subsequent time periods; 2) we consider a set of prerequisite commodities for each commodity at a node and add the corresponding precedence constraint for the relationship. We consider a Constraint and Column Generation (CCG) method for solving the MIP R-MC-MVDFF by decomposing the model into path-finding subproblems and a main problem maximizing value over the worst case. We investigate the efficiency of our solution methods through rigorous computational results.

MA28

Building B Level 2 B202

Advanced Optimization for Mixed Integer Programming

Invited Session

Computing Society

Chair: Mikhail Bragin, Southern California Edison, Marina del Rey, CA, United States

1 - Lp Sampling for Improved Milp Analysis and Solution**John Chinneck, Carleton University, Ottawa, ON, Canada**

A new heuristic that generates and solves random LPs derived from a given MILP can be helpful in two situations. (1) If the MILP is infeasible, it helps to find a minimal set of integer restrictions (IRs) causing infeasibility. (2) If the MILP is feasible, it can help find an initial MILP-feasible or near-feasible solution. In case (1) most solvers do not attempt to isolate a minimal set of IRs because it takes a very long time, though this information is useful to the analyst. Doing so can also greatly speed up the isolation of the entire infeasibility (including linear constraints and variable bounds). The heuristic treats all variables as continuous and solves several LPs with varying objective function coefficients for the nominally integer variables. This is often surprisingly useful in quickly reducing the original model in case (1) and quickly providing a MILP-feasible or near-feasible solution in case (2). Experimental results are provided

2 - Learning to Tune Parameters in Multi-Period Blending Problem**Asha Ramanujam, Purdue University, West Lafayette, IN, United States, Dimitri Papageorgiou, Can Li**

The multi-period blending problem, also known as the multi-period pooling problem, involves blending streams in blenders to obtain products across different time periods. Optimizing the blending system involves solving an NP-hard non-convex mixed integer quadratically constrained problem (MIQCP). Large-scale versions of these MIQCPs can be very difficult to solve to provable optimality and even for high-quality primal solutions. In this talk, we introduce a novel, scalable method that involves tuning parameters in a low-fidelity approximation of the MIQCP so that it can achieve similar performance as the high-fidelity full-space MIQCP.

3 - Acceleration of Level Adjustment for the Polyak Stepsize: Applications to Mixed Integer Programming and Machine Learning**Mikhail Bragin, Southern California Edison, Marina del Rey, CA, United States**

The Polyak stepsize has been widely used for convex problems. However, the stepsize requires a generally unknown optimal value. To guarantee convergence, we develop a novel level adjustment approach to successively estimate the optimal value with provable convergence. This is achieved by developing a decision-based procedure through a novel easy-to-solve "Polyak Stepsize Violation Detector" (PSVD) linear constraint satisfaction problem. Once a violation is detected, a tighter level value is obtained. Furthermore, because subgradients may be computationally costly, we develop an approximate subgradient method. Through a series of empirical tests of convex functions with diverse characteristics, we illustrate the practical advantages of our approach as compared to existing methods. For Mixed-Integer Programming (MIP) problems, our integration of Lagrangian Relaxation with the Polyak stepsize demonstrates significant efficiency gains, optimizing Lagrangian dual problems in operations research, power systems, transportation, and manufacturing sectors leading to speedups of 2-3 orders of magnitude as compared to standard MIP methods.

4 - Screening Cut Generation for Sparse Ridge Regression**Guanyi Wang, National University of Singapore, Singapore, Singapore**

Sparse ridge regression is widely utilized in modern data analysis and machine learning. However, computing globally optimal solutions for sparse ridge regression is challenging, particularly when samples are arbitrarily given or generated under weak modeling assumptions. This paper proposes a novel cut-generation method, Screening Cut Generation (SCG), to eliminate non-optimal solutions for arbitrarily given samples. In contrast to recent safe variable screening approaches, SCG offers superior screening capability by identifying whether a specific $\{\pm 1\}$ combination of multiple features (binaries) lies in the set of optimal solutions. This identification is based on a convex relaxation solution rather than directly solving the original sparse ridge regression. Hence, the cuts generated by SCG can be applied in the pre-processing step of branch-and-bound and its variants to construct safe outer approximations of the optimal solution set. Numerical experiments are reported to validate the theoretical results and demonstrate the efficiency of SCG, particularly in hard real instances and synthetic instances with high dimensions, low ridge regularization parameters, or challenging modeling assumptions.

5 - Distributed Quantum Computing for Fast Restoration of Distribution Networks with Grid-Forming Ibrs**Peng Zhang, Stony Brook University, New York, NY, United States**

We will present a distributed quantum computing (DQC) framework for scalable restoration of distribution networks with grid-forming IBRs. By decomposing the global restoration problem into independent spatial subproblems, the DQC restoration framework enables parallel execution on heterogeneous quantum resources and significantly accelerates restoration decision-making, leading to reduced load loss and improved system reliability. A quantum restoration formulation that incorporates droop constraints of grid-forming IBRs is introduced to maintain frequency and voltage security throughout the restoration process. To manage the increasing model complexity from temporal decision variables, the time-series MILP models of spatial subproblems are further decomposed into smaller temporal subproblems, facilitating scalable implementation on NISQ devices. A compact and lossless transformation from MILP to QUBO is designed through a set of encoding, scaling, and tuning techniques. This minimizes the number of required qubits while preserving solution fidelity, allowing for efficient mapping of service restoration problems onto today's quantum annealers and gate-based devices.

6 - Quantum Subspace Decomposition for Large-Scale Power Grid Analysis**Yifan Zhou, Stony Brook University, Stony Brook, NY, United States**

Quantum computing holds immense potential for resolving computationally intractable problems, which ignites hope for a "quantum leap" in power grid computing. However, developing scalable quantum algorithms for power grid analysis remains challenging, due to difficulties in designing practically implementable quantum circuits and maintaining robustness under varying grid conditions and quantum noise.

This talk will present our recent work on advancing large-scale, quantum-enabled power grid analysis using Quantum Krylov Subspace (QKS) methods. The core idea is to project high-dimensional power grid models onto a lower-dimensional quantum subspace through a series of computationally manageable quantum circuits and small-scale classical problems. This approach allows solving large-scale problems while avoiding the need for complex, high-depth circuits (e.g., HHL) or heuristic variational optimization techniques. It will offer a practical and scalable path toward quantum-enhanced power grid computation for systems with tens of thousands of buses on near-term quantum devices.

MA29

Building B Level 2 B203

Algorithms for Grid Integration of AI

Invited Session

ENRE: Electricity

Chair: Vladimir Dvorkin, University of Michigan, Ann Arbor, MI, United States

1 - Toward Energy-Optimal AI Systems**Mosharaf Chowdhury, University of Michigan, Ann Arbor, Ann Arbor, MI, United States**

Generative AI adoption and its energy consumption are skyrocketing. For instance, training GPT-3, a precursor to ChatGPT, consumed an estimated 1.3 GWh of electricity in 2020. By 2022, Amazon trained a large language model (LLM) that consumed 11.9 GWh, enough to power over a thousand U.S. households for a year. AI inference consumes even more energy, because a model trained once serves millions. This surge has broad implications. First, energy-intensive AI workloads inflate carbon offsetting costs for entities with Net Zero commitments. Second, power delivery is now the gating factor toward building new AI supercomputers. Finally, this hinders

deploying AI services in places without high-capacity electricity grids, leading to inequitable access to AI services. In this talk, I will introduce the ML Energy Initiative, our effort to understand AI's energy consumption and build a sustainable future by curtailing AI's runaway energy demands. I will introduce tools to precisely measure AI's energy consumption and findings from using them on open-weights models, algorithms to find and navigate the Pareto frontier of AI's energy consumption, and the tradeoff between performance and energy consumption during model training. I will also touch upon our solutions to make AI systems failure-resilient to reduce energy waste from idling. This talk is a call to arms to collaboratively build energy-optimal AI systems for a sustainable and equitable future.

2 - Learning-Augmented Competitive Algorithms for Spatiotemporal Online Allocation with Deadline Constraints

Adam Lechowicz, University of Massachusetts Amherst, Amherst, MA, MA, United States

We introduce and study spatiotemporal online allocation with deadline constraints (SOAD), a new online problem motivated by emerging challenges in sustainability and energy. In SOAD, an online player completes a workload by allocating and scheduling it on the points of a metric space (X, d) while subject to a deadline T . At each time step, a service cost function is revealed that represents the cost of servicing the workload at each point, and the player must irrevocably decide the current allocation of work to points. Whenever the player moves this allocation, they incur a movement cost defined by the distance metric $d(\cdot, \cdot)$ that captures, e.g., an overhead cost. SOAD formalizes the open problem of combining general metrics and deadline constraints in the online algorithms literature, unifying problems such as metrical task systems and online search. We propose a competitive algorithm for SOAD along with a matching lower bound establishing its optimality. Our main algorithm, ST-CLIP, is a learning-augmented algorithm that takes advantage of predictions (e.g., forecasts of relevant costs) and achieves an optimal consistency-robustness trade-off. We evaluate our proposed algorithms in a simulated case study of carbon-aware spatiotemporal workload management, an application in sustainable computing that schedules a delay-tolerant batch compute job on a distributed network of data centers. In these experiments, we show that ST-CLIP substantially improves on heuristic baseline methods.

3 - The Dynamical Profile and Impact of Large Language Model Power Consumption

Yuzhuo Li, University of Alberta, Edmonton, AB, Canada

Large language models (LLMs) have driven unprecedented investment in AI infrastructure, with multi-hundred-billion-dollar data center expansions planned over the next 3-5 years. However, the unique dynamical power consumption profile of LLMs poses significant challenges for power grid integration and reliability. Unlike traditional loads, AI infrastructure exhibits ultra-low inertia, sharp power surges and dips, and substantial peak-to-idle power ratios spanning from hundreds of watts to gigawatts. These transient dynamics create never-before-seen grid integration challenges. This work examines the scale and characteristics of LLM power consumption, analyzing transient behavior across various operational scenarios to characterize the complex patterns of AI workload dynamics. We develop preliminary mathematical frameworks to capture key aspects of this behavior and assess their multifaceted impact on existing power grids. Our analysis reveals critical reliability and resilience concerns that emerge from the disruptive nature of AI loads, including rapid power fluctuations that challenge traditional grid stability mechanisms. The findings highlight the urgent need for interdisciplinary approaches combining machine learning expertise with power systems engineering to address these integration challenges. As AI technologies continue evolving rapidly, characterizing and understanding their dynamical power profiles becomes essential for sustainable infrastructure development. This research provides foundational insights for developing algorithms and strategies for effective grid integration of AI, offering a starting point for researchers and practitioners working at the intersection of artificial intelligence and power systems.

4 - Carbon-Aware Computing for Data Centers with Probabilistic Performance Guarantees

Sophie Hall, ETH Zürich, Zürich, Switzerland, Francesco Micheli, Giuseppe Belgioioso, Ana Radovanovic, Florian Dörfler

Data centers are significant contributors to carbon emissions and can strain power systems due to their high electricity consumption. To mitigate this impact and to participate in demand response programs, cloud computing companies strive to balance and optimize operations across their global fleets by making strategic decisions about when and where to place compute jobs for execution. In this work, we introduce a load shaping scheme which reacts to time-varying grid signals by leveraging both temporal and spatial flexibility of compute jobs to provide risk-aware management guidelines and job placement with provable performance guarantees based on distributionally robust optimization. Our approach divides the problem into two key components: (i) day-ahead planning, which generates an optimal scheduling strategy based on historical load data, and (ii) real-time job placement and (time) scheduling, which dynamically tracks the optimal strategy generated in (i). Our optimal scheduling strategy allows for incorporation of grid constraints such as power limits and ramping constraints. Further, we use a stochastic model to handle demand response events which ensures job completion requirements still hold in expectation. We validate our method in simulation using normalized load profiles from randomly selected Google clusters, using real carbon load signals, and incorporating ramping limits. We demonstrate significant reductions in carbon cost and peak power with our approach compared to myopic greedy policies, while maintaining computational efficiency and abiding by system and grid constraints.

5 - Carbon-Intelligent Compute Management Leveraging Temporal and Spatial Flexibility for Enhanced Emission Reduction

Utkarsh Shah, Google, Inc, Seattle, WA, United States, Bokan Chen, Ana Radovanovic

The carbon dioxide emitted per kilowatt-hour on an electricity grid varies based on the time of day and significantly by location due to different power generation methods. Google employs a Carbon-Intelligent Compute Management system that leverages temporal flexibility to decrease carbon emissions from electricity consumption by adjusting the timing of adaptable workloads. Expanding on this, this work introduces geographical relocation of these workloads, adding spatial flexibility. The practicality of this relocation depends on factors like data dependencies, network availability, and regulations. A spatial flexibility classifier (SFC) evaluates a job's spatial flexibility, which then informs both offline demand forecasting and online real-time cluster selection systems. Offline elements, including day-ahead spatial demand prediction, carbon intensity forecasting, and a risk-aware optimization model, produce virtual capacity curves (VCCs) for Google's data centers. These VCCs control capacity distribution across different clusters and times, ensuring the daily completion of all flexible workloads across the entire fleet with high reliability. Google's operational data shows that VCCs, in conjunction with the SFC, effectively manage cluster capacity by shifting flexible workload execution to locations and times with lower carbon intensity when a cluster's energy supply has a high carbon footprint.

MA30

Building B Level 2 B204

Grid Modernization Investment and Operational Security

Invited Session

ENRE: Electricity

Chair: Satender Gunwal, University of Illinois Chicago, CHICAGO, 60607, United States

Co-Chair: Selva Nadarajah, University of Illinois Chicago, Downers Grove, IL, United States

Co-Chair: Todd Levin, Argonne National Laboratory, Lemont, IL, 60439, United States

1 - Modernizing Expansion Planning and Probabilistic Resource Adequacy Modeling for Spatial Reliability: Metrics and Machine Learning Feedback

Satender Gunwal, University of Illinois Chicago, CHICAGO, IL, United States, Selva Nadarajah, Todd Levin

Generation and Transmission Expansion Planning (GTEP) is essential for ensuring adequate investments in electricity supply and transmission to meet growing demand, evolving system needs, and changing technology costs. Traditional GTEP models typically focus on system-wide cost minimization and reliability via deterministic optimization. Subsequently, Probabilistic Resource Adequacy Models (PRAM) simulate operational reliability of the resulting system. However, improvements in traditional system-wide reliability metrics do not guarantee spatial reliability—consistent electricity supply across different regions—leading to critical spatial disparities. Existing approaches to spatial reliability often result in computationally intractable stochastic models or rely on ad hoc integrations between GTEP and PRAM, highlighting structural limitations. As ratepayers ultimately fund network expansions, addressing spatial disparities is essential to improve societal outcomes, build public trust, streamline permitting, and support policy objectives. Increased renewable intermittency further amplifies these challenges, necessitating modernized planning approaches.

This paper proposes the use of economic fairness metrics based on unserved energy to assess and improve spatial reliability, particularly benefiting underserved regions and vulnerable consumers. Using the county-level ERCOT network, we demonstrate that traditional GTEP models achieve aggregate reliability improvements yet fail

spatially. We then introduce a machine learning-based feedback mechanism that links GTEP and PRAM efficiently through MIP-representable random feature mapping with rectified linear unit (ReLU) activation. This approach offers computational efficiency and interpretability without disruptive software infrastructure changes. We illustrate the benefits of this feedback mechanism using Garver 6-bus test system to demonstrate its effectiveness and interpretability.

2 - Decomposition Algorithms for Stochastic Capacity Expansion with Endogenous Adequacy Targets

Marilena Zampara, National Technical University of Athens, Athens, Greece, Daniel Felipe Avila, Anthony Papavasiliou

The scope of this work is to test alternative algorithmic schemes for addressing a large stochastic capacity expansion problem with specific constraints on expected energy not served. The challenge lies in decoupling scenario-specific variables, which are originally bundled by adequacy constraints, so that scenario-specific problems can be solved in parallel. The examined schemes include Benders' decomposition, the projected subgradient method, the level method, and Dantzig-Wolfe decomposition. The schemes are tested in a pan-European capacity expansion model, considering 35 scenarios for climatic conditions.

3 - Reliability Impacts of Energy Storage Profitability Considerations in Electricity Market Models

Mark Noll, Northwestern University, Evanston, IL, United States, David Morton

Many options exist for scheduling energy storage in resource adequacy models. An approach that has become widespread in recent work has been to establish a day-ahead schedule for storage and then assume that operations in real-time do not deviate from the day-ahead schedule except in extreme circumstances, such as to replace operating reserves or prevent load shedding. In reality, storage units that operate in restructured electricity markets have more leverage to choose their schedules in the real-time market than assumed by this modeling paradigm. From a reliability perspective, there is a concern that storage units may prematurely deplete their stored energy when RT prices are high but storage is not needed to prevent a loss-of-load event, leaving the storage fleet without stored energy to prevent a loss of load event later in the day.

In this work, we explore the conditions under which storage units, operating with imperfect foresight in day-ahead and real-time markets, may choose to deviate from their day-ahead schedule in real-time operations. To represent the perspective of storage units operating under uncertainty, we solve a sequence of hourly rolling horizon stochastic optimization models that maximize risk-adjusted future profits for the storage unit, accounting for the storage unit's day-ahead schedule and forecasts of real-time prices in the future. We aim to clarify the conditions under which profit-seeking storage behavior may reduce availability during reliability events. These insights can inform future RA modeling assumptions and operational practices in modern electricity markets.

4 - Resilient Power System Design Against Climate Hazards

Brayam Valqui Ordenez, EPRI, Palo Alto, CA, United States, Miguel Ortega-Vazquez, Eknath Vittal

Developed and applied a resilient generation expansion model to the Alliant bulk power system to evaluate long-term investment strategies under climate-driven extreme events. Resilience is embedded into capacity the planning model by explicitly representing extreme event risks, which are multifaceted and uncertain. The planning approach optimizes expansion and hardening investments by treating events as multicriteria stress tests that prevent unacceptable unmet demand and damage outcomes, while costs are optimized over ordinary operating conditions. Each event contains multiple hazards that disrupt operations, damage assets (as a function of their vulnerabilities), and increase demand. System resilience is measured using a load loss tolerance (LLT) metric, which defines the allowable level of unmet demand during extreme conditions. Resiliency is guaranteed by sensitivity analysis across a range of LLT values while also enforcing a net-zero emissions target by 2050. Results show that lower LLT thresholds shift the optimal portfolio toward firm, dispatchable, carbon-free capacity, with increased investment in technologies such as hydrogen-enabled RICE and gas peaker RICE, as well as targeted hardening of thermal generators through winterization and chiller technologies. These shifts reflect the dual objectives of enhancing resilience and achieving long-term decarbonization. At the same time, investments in wind, solar, and storage decline moderately as the system prioritizes clean dispatchable capacity under stress conditions. These changes result in total investment and operational cost increases ranging from 0.2% (LLT0.3) to 11% (LLT0) relative to the baseline scenario (i.e., full loss during stress scenarios).

MA31

Building B Level 2 B205

Improving Mine Planning and the Minerals Supply Chain, and mitigating wildfires

Invited Session

ENRE: Natural Resources

Chair: Alexandra Newman, Colorado School of Mines, Golden, CO, United States

1 - Shortest Path: Ramp Optimization in Open Pit Mining

Felix Ayaburi, Colorado School of Mines, Lakewood, CO, United States, Bingshen Lu

Haul Road Optimization in Open-Pit Mining aims to reduce fuel consumption and operational costs by improving ramp design for large haul trucks. Traditional methods rely on heuristics and often overlook the impact of terrain, slope, and truck constraints on efficiency.

We model the mine surface as a 3D graph where nodes are connected based on distance and slope thresholds. Using Python and NetworkX, we implement the A* algorithm with edge weights that penalize steep gradients and tight turns, accounting for vehicle size, turn radius, and ramp width. This allows us to generate realistic, fuel-efficient haul routes for Komatsu 930Es and Hitachi EH5000ACs.

Current results demonstrate optimized surface paths that reduce travel distance and fuel usage. Future work will extend to subsurface planning, evaluating the trade-offs between material removal and long-term haulage efficiency.

2 - Multi-Objective Optimization of Drill Pattern Design for Increased Ore Recovery and Reduced Impact Mining

Ozan Perincek, Rio Tinto, Perth, Australia, Ryan Loxton, Sarang Kulkarni, Daniel Arthur

Drill and blast processes aim to optimize the fragmentation of rocks for efficient hauling and crushing. Due to environmental considerations and scarcity of resources, the focus is now shifting towards reducing carbon emissions and minimizing ore loss. Blasting is more energy efficient than crushing for breaking rocks and controlling rock movement during blasting can reduce dilution and improve ore recovery. Existing work has used initiation timing design to blast ore away from waste but does not consider the optimal placement of drill holes to enhance ore-waste separation. Poor drill hole placement may lead to suboptimal fragmentation with large rock sizes, resulting in energy-intensive downstream processing. We consider the multi-objective optimization of energy consumption and ore-waste separation by determining drill hole locations and explosive product in each hole. In a two-phase approach, first, we minimize the total mining energy by achieving optimal fragmentation. Second, using this energy as a bound, we re-optimize the positions of the drill holes to maximize ore recovery. The algorithm uses a heuristic approach to deal with complexities such as nonlinearity and an unknown number of drill holes to provide decision support for the fragmentation vs ore recovery tradeoff. Additionally, the model considers ease of navigation within the drill pattern for drills and explosive delivery trucks. Simulations of six blast scenarios using data from a Rio Tinto mine site in Western Australia indicate that trading off fragmentation and optimizing hole position with specialized initiation timing can recover between 3% to 9% more ore than just utilizing timing.

3 - Optimizing Haul Trucks Maintenance Scheduling in Open Pit Mines: Case Study on Modelling Around Complexity

Fredrick Ehichioya, Imperial Oil, Calgary, AB, Canada, Bart Maciszewski, Siddhesh Inamdar, Jose Santiago Rodriguez Osorio, Anuraj Grewal, Christine Vilijoen, Amber McKay, Ervin Balleza, Andrew Milne, Jacob Olafsen, Cathy Giang, Hussam Mansour, Pujit Jetley, Brendan Cumming

We present a case study of building a model to schedule maintenance for heavy haul trucks in an oil sands mine and its refinement over time to expand scope and meet business needs. In this problem the trucks require periodic preventive maintenance based on operating hours and in addition to break-in work. There is a need to schedule

close to due dates while grouping activities to minimize the number of truck visits to the shop. Activities need to be performed on compatible bays and need to respect the amount of available resources. We decompose the problem into a grouping and detailed scheduling stage and solve it using constraint programming with a rolling horizon approach. We show how we adapted the model to cover diverse mine equipment fleets and complex business requirements through separation of data processing, mathematical modelling, and integration of machine learning.

4 - Use of Gen AI to Improve Mining Workforce Efficiency

Hannah Zhao, Imperial Oil Limited, Calgary, AB, Canada, Amy Liang, Fredrick Ehichioya, Trevor Lew, Anuraj Grewal, Bryan Kan

At UDA, we are building applications that leverage generative AI to enhance the efficiency of troubleshooting processes in mining operations. By providing real-time information retrieval, the AI system optimizes troubleshooting workflows and improves worker productivity.

MA32

Building B Level 2 B206

Electricity and Environmental Market Interactions

Invited Session

ENRE: Environment and Sustainability

Chair: Christoph Graf, New York University, New York., United States

Co-Chair: Dharik Mallapragada, New York University, Brooklyn, NY, 11201, United States

1 - Carbon-Aware Electric Load Shifting and Its True Impact on Emissions

Line Roald, University of Wisconsin - Madison, Madison, WI, United States, Joseph Gorka, Wenqian Jiang, Noah Rhodes

An increasing number of individuals, companies and organizations are interested in computing and minimizing the carbon emissions associated with their real-time electricity consumption. To achieve this, they require a carbon signal, i.e. a metric that defines the real-time carbon intensity of their electricity supply. Unfortunately, in a grid with multiple generation sources and multiple consumers, there are multiple ways to define this signal, each of which has a distinct set of properties and method of calculation. It remains unclear which signal best quantifies the carbon footprint of electricity. This paper seeks to inform the discussion about which carbon signal is better or more suitable for two important use cases, namely carbon-informed load shifting and carbon accounting. We do this by developing a new software package `ElectricityEmissions.jl`, that computes several established and newly proposed carbon emission metrics for standard electric grid test cases. We also demonstrate how the package can be used to investigate the effects of using these metrics to guide load shifting. Our results affirm previous research, which showed that the choice of carbon emission metric has significant impact on shifting results and associated carbon emission reductions. In addition, we demonstrate the impact of load shifting on both the consumers that perform the shifting and consumers that do not. Disconcertingly, we observe that shifting according to common metrics such as average carbon emissions can reduce the amount of emissions allocated to the consumer doing the shifting, while increasing the total emissions of the power system.

2 - Impacts of Varying Approaches to Clean Electricity Accounting

Pieter Gagnon, Colorado School of Mines, Golden, CO, United States

What is the system-level impact of different approaches for REC-based clean electricity accounting? Does lower participation in a program necessarily mean the program has lower impact, or can more stringent accounting approaches drive greater change with lower participation? In this study we use a capacity expansion model to explore the interaction of electricity and environmental markets in the US, and how design attributes like spatial and temporal matching can influence the impact of an accounting approach. This work explores a range of assumptions about the elasticity of consumer's demand for clean power, to characterize how changes in the stringency of clean electricity accounting approaches can influence their ultimate impact.

3 - Market Signals and Incentives Aren't Sufficient: the Case for Supply-Chain Modeling for Grid Expansion

Boyu Yao, Johns Hopkins University, Baltimore, MD, United States, Yury Dvorkin

Until recently, grid expansion planning has assumed that asset deployment responds coherently to both electricity and environmental market signals, while overlooking upstream constraints. In practice, however, supply chain dynamics—such as material shortages, spatial constraints, and deployment timelines—distort the link between market signals and actual deployment. To be effective, grid expansion must account for supply chain and infrastructure realities to ensure that market signals translate into achievable outcomes. The first part of this talk evaluates renewable capacity expansion under realistic constraints, including material availability, land use limits, and lead times. Results show that meeting clean energy targets with renewables alone may be infeasible once rising demand from electric vehicles, data centers, and other sectors are taken into account. The second part extends the analysis to the demand for solid-state electronics—including inverters, converters, and transformers—which are critical for both integrating renewables and replacing aging grid infrastructure. Projections of future infrastructure needs highlight substantial material and component requirements, alongside growing cross-sectoral competition for these resources. Together, these findings point to a structural misalignment between market-based incentives and the physical feasibility of grid asset deployment.

4 - Perspectives on the System Impacts of Alternative Clean Electricity Procurement Strategies Using An Investment Planning Framework

Dharik Mallapragada, Tandon School of Engineering, New York University, Brooklyn, NY, United States, Zhetao Chen, Christoph Graf

Quantifying CO₂ emissions from electricity use is increasingly important for economy-wide decarbonization, as consumers seek to evaluate the impact of actions like electrification and clean power procurement. This has led to growing interest in granular emissions accounting methods that capture dynamics of grid operations to inform consumer procurement strategies. Two emerging approaches include: (i) *Emissions matching*, where consumers aim to minimize their grid emissions, measured using location- and time-specific marginal emissions factors; and (ii) *Temporal matching*, which aligns electricity use with low-carbon generation (e.g., hourly) as a proxy for emissions tracking.

Capacity expansion models (CEM) have proven valuable in systematically assessing the long-term cost and emissions implications of consumer-driven clean electricity procurement under varying grid and technology scenarios. We review CEM-based assessments of clean electricity procurement to highlight areas for further research: (a) impacts of inter-annual weather variability on matching performance; (b) interactions—competitive or coordinated—among consumers with differing strategies, load shapes, and locations; and (c) the influence of diverse consumer preferences around risk and clean energy procurement.

We then explore a decentralized framework for long-term investment planning in the power sector with the potential to address some of the above gaps around consumer coordination and preferences. This framework employs the Alternating Direction Method of Multipliers (ADMM) to compute equilibria in a non-cooperative game among heterogeneous consumers. Using a Texas case study, we compare the outcomes and scalability of the ADMM-based framework with those using a traditional CEM, offering insights into the trade-offs between centralized and decentralized methods.

5 - Optimizing the Economics and Reliability of 24/7 Power Purchase Agreements with Hybrid Energy Resources

Zhi Zhou, Argonne National Laboratory, Lemont, IL, United States, Dongwei Zhao

Power Purchase Agreements (PPAs) are increasingly favored by large energy consumers, such as AI data centers, for procuring reliable and cost-effective electricity. While traditional PPAs typically operate on a pay-as-produced basis with annual supply-demand balancing, emerging needs are driving a shift toward hourly matching of

supply and demand. Achieving this level of granularity requires the integration of flexible energy resources, such as hydropower and energy storage, to complement variable renewables like wind and solar. These hybrid configurations form the foundation of 24/7 flexible energy PPAs.

This study presents an integrated optimization framework for designing 24/7 PPA from the PPA provider's perspective. The provider procures hybrid resources and designs PPAs to meet users' hourly electricity demand. The framework incorporates both dispatchable resources (e.g. reservoir hydropower) and non-dispatchable renewables to enhance overall supply flexibility and economics. It jointly optimizes resource procurement, operational scheduling, and PPA pricing, with the objective of minimizing contract costs or maximizing provider profit. Simulation results demonstrate the critical role of flexible resources in enabling cost-effective 24/7 energy delivery. Sensitivity analyses further illustrate how hydropower flexibility and pricing mechanisms affect revenue distribution and cost allocation among providers, users, and the broader energy system.

MA33

Building B Level 2 B207

Applications of OR/IE Models in Health and Public Policy

Invited Session

Health Applications Society

Chair: Suyanpeng Zhang, Northwestern University, Evanston, IL, United States

1 - Improving Cystic Fibrosis Screening Through a Novel Testing Design

Hussein El Hajj, Santa Clara University, Santa Clara, CA, United States, Douglas Bish, Ebru Bish

Newborn screening (NBS) plays a key role in detecting life-threatening genetic disorders, with cystic fibrosis (CF) being a prominent example. With over 1,000 CF-causing variant types, a genetic test that can detect enough variant types for sufficient screening accuracy is expensive and capacity-constrained. Consequently, in the US, the first tier of testing uses an inexpensive biochemical test that measures immunoreactive trypsinogen (IRT) levels, which are often elevated in CF cases. Genetic testing is reserved for a small fraction of newborns with elevated IRT levels, based on an IRT threshold. However, setting an optimal threshold presents challenges, particularly due to variations in IRT levels and CF prevalence among different racial and ethnic groups. As a result, the current practice of using a single IRT threshold for all newborns is not necessarily the most efficient or equitable strategy.

We introduce a novel screening design, integrating an inexpensive small-panel genetic test (DNA) before the IRT test. This strategy customizes IRT thresholds based on the number of variants detected by the small-panel DNA. This design leads to novel optimization problems, including variant type selection for the small-panel DNA and IRT threshold customization. We establish key structural properties of optimal IRT thresholds and develop Pareto frontiers that inform decision-makers about the efficiency versus equity trade-off, considering a minimax type equity measure. The case study, utilizing CF NBS data from the states of New York and California, demonstrates the potential efficiency and equity gains of this novel approach compared to current practice and race-based benchmarks.

2 - Examining the Unintended Impacts of Disruptions to the Illicit Drug Market on Violent Crime and Overdose Death

Regan Richardson, North Carolina State University, Raleigh, NC, United States, Laura Mora, Osman Ozaltin, Maria Mayorga, Brad Ray, Amang Sukasih, Phillip Lee, Lance Couzens, Nur Kaynar

The use of illicit substances remains an ongoing and widespread part of society, with the illegal drug market estimated to be 1% of the total global trade. The true scale of this market is unknown because it is an illicit underground trade that is poorly understood, dynamic, and complex. While there have been trends towards decriminalization or legalization of marijuana in the United States, the overarching policy and strategy has been consistently been to allocate large amounts of resources to law enforcement agencies to disrupt and dismantle illicit distribution networks. Therefore, the purpose of this project is to utilize national integrated data to develop an understanding of the impact of law enforcement disruptions to the illicit drug supply on both public health (e.g. overdoses) and public safety (e.g. violent crime) outcomes. In the present study, we utilize multivariate granger-causality to investigate the relationships between police drug seizures and violent crime rates, and between police drug seizures and overdose death rates using data from the National Incident Reporting System (NIBRS) and the National Center for Health Statistics Mortality Data from CDC WONDER.

3 - Multi-Modal Machine Learning for Breast Cancer Recurrence Prediction

Jiahao Shao, The University of Tennessee Knoxville, Knoxville, TN, United States, Xudong Wang, John Bell, Xueping Li, Bing Yao

Breast cancer recurrence, a leading cause of long-term mortality among survivors, requires timely and accurate risk assessment to guide follow-up care and treatment planning. Traditional predictive models, often limited to either structured or unstructured data alone, struggle to capture the full clinical context. This study examines the impact of integrating multi-modal clinical data, including treatment records, pathology reports, and clinician notes, on recurrence prediction. Large language models (LLMs) are used to extract key prognostic variables such as tumor grade and hormone receptor status from free-text pathology narratives. These outputs are cross-validated against rule-based extractions to improve consistency and build trust in downstream modeling. We also benchmark performance against commonly used feature sets from prior breast cancer studies to assess the added value of multi-modal integration. Single-source and multi-modal inputs are evaluated across a range of machine learning models, including both interpretable and black-box approaches. Results show that multi-modal integration consistently improves predictive accuracy compared to single-modal methods.

4 - Deepsynergy: Fusing Hybrid Optimization and Shallow-Deep Ensembles for Enhanced Breast Cancer Survival Analysis

Nikoo Ebrahimi, Binghamton University (SUNY), Binghamton, NY, United States, Sadaf Tabatabaee, Sreenath Chalil Madathil, Mohammad Khasawneh

Accurate breast cancer survival prediction is a critical clinical challenge in machine learning. We introduce a deep learning framework, called DeepSynergy, that combines shallow network insights and hybrid optimization to predict breast cancer survivability using the SEER breast cancer dataset. This dataset provides comprehensive, de-identified records of female patients diagnosed with complete follow-up through a defined cutoff period. This research incorporated six variants and three synergistic models for the base DeepSurv model to enhance training and predictive performance. The six variants include checkpointing, gradient clipping, Optuna-based hyperparameter tuning, refined best hyperparameter with AdamW optimization, stochastic gradient descent, and a hybrid optimizer combining SGD warm-up with AdamW fine-tuning. The synergistic model includes augmented deep model inputs with intermediate hidden representations, a gated-fusion ensemble using the shallow log-risk score, and a parallel dual-path architecture for multimodal representation learning. We applied inverse probability of censoring weighting (IPCW) to mitigate censoring bias and improve reliability. All models were evaluated using C-index, IBS, and time-dependent AUC at 1, 3, 5, and 10 years. The hybrid optimizer and the ensemble model demonstrated superior calibration, faster convergence in risk discrimination, and strong early-stage survivor stratification, highlighting their potential for clinical decision-making in early phases of treatment. We employed SHAP and LIME for post hoc interpretability to improve model transparency. Both methods consistently identified nodal involvement, tumor grade, and staging completeness as the most influential predictors across models. These findings underscore the potential of DeepSynergy as an explainable and deployable decision support tool in oncology.

5 - Personalized Risk Assessment and Treatment Recommendation for Substance use Disorders

Eric Pulick, University of Wisconsin - Madison, Madison, WI, United States, Yonatan Mintz

Substance use disorders (SUDs) are a widespread and growing global health concern. In the United States (US) alone, nearly fifty million people meet the diagnostic thresholds for an SUD. However, fewer than 15% of those with SUDs receive any form of treatment. Even fewer receive the type of long-term, continuing care often associated with successful recovery. Digital therapeutics platforms (e.g., an app on a mobile device) are emerging as a low-cost and low-burden way to interact with SUD patients, both to gather longitudinal data and curate resources for continuing care. In this talk, we discuss the use of personalized state space models for assessing SUD patient risk and their extension to adaptive control frameworks for personalized treatment selection. We validate these approaches using datasets built from digital therapeutics studies that followed patients with alcohol- and opioid use disorders for 3-12 months.

6 - Representing Caregiver Burden in Observational Studies: Development of the Caregiver Burden Index (Carebi)

Forough Mahpouya, University at Buffalo, Buffalo, NY, United States, Sabrina Casucci, Suzanne Sullivan, Christopher Barrick

Informal caregiving, provided by over 53 million adults in the US, often carries a significant emotional, physical, and financial toll, or caregiving burden. Unpaid caregivers manage responsibilities that significantly affect both their own well-being and their care recipients' outcomes, yet caregiver burden is rarely reflected in healthcare decision-making frameworks. Existing caregiver burden instruments, while valuable in clinical research, often lack compatibility with observational datasets regularly used in health services research and planning. This study introduces the Caregiver Burden Index (CareBI) using data from the National Study of Caregiving, that can be used to represent caregiver burden in quantitative models and research based on observational data. We will discuss our development process, relevant indicators and burden domains, and alignment of CareBI with the Zarit Burden Index, a widely used tool for prospectively assessing caregiver burden. We will also discuss how we evaluated our scale for construct validity by assessing its relationship with caregiver and care recipient outcomes, as well as sensitivity to known burden-related risk and mitigation factors. Early findings affirm the scale's utility in categorizing low-, moderate-, and high-burden caregivers and guiding resource-oriented strategies. CareBI represents a reproducible tool for embedding caregiver metrics into health operations, predictive modeling, and public policy frameworks, and provides a template for applying operations research and industrial engineering methods to psychosocial measurement challenges in aging and long-term care.

7 - Screening and Sensibility: a Societal Approach to Disease Screening for Patient with Multiple Chronic Conditions

Ali Hajjar, MBSC, King Abdullah Economic City, Saudi Arabia, Oguzhan Alagoz

Existing clinical practice guidelines tend to overlook the unique requirements of patients with multiple chronic conditions (MCC) by primarily focusing on the management of individual diseases. This inadequately personalized approach leads to adverse events and places a financial strain on the healthcare system, particularly in countries with large or growing populations. To address this issue, our study utilizes a stochastic modeling approach to optimize disease screening decisions while considering financial and resource limitations from a broader societal perspective.

8 - Data-Driven Frameworks for Quantifying Relationships Between Diseases

Hari Balasubramanian, University of Massachusetts, Amherst, Amherst, MA, United States

The prevalence of multiple chronic diseases among patients is increasing globally. Systematic approaches to identify relationships between diseases are essential for: (1) prevention, diagnosis, and treatment planning; and (2) epidemiological insights on how different organs and body systems might be linked together. In this talk, we discuss examples from the literature, methodologies, and datasets that can help identify associations between diseases.

MA34

Building B Level 2 B208

Recent Advances in Mixed-Integer and Global Optimization

Invited Session

OPT: Global Optimization

Chair: Siva Ramani, Rice University, Houston, TX, United States

Co-Chair: Rohit Kannan, Virginia Tech, Blacksburg, VA, United States

1 - Water Network Design and Operation: Leveraging Linear Approximations

Sourabh Kumar Choudhary, Georgia Institute of Technology, Atlanta, GA, United States, Santanu Dey, Nick Sahinidis

This study concerns the optimal design and operation of produced water and urban water networks. We address a key challenge in solving these: the representation of frictional pressure changes across network nodes using nonlinear constraints, typically modeled by the Hazen-Williams equation. For the optimization of produced water networks, we analytically show the effectiveness of using a standard piecewise linear approximation to generate near-optimal solutions for the original problem for a general network. Computational results with real-world problems confirm the success of this approach. However, in the context of urban water network design problems, we recognize the limitations of a direct piecewise-linear approximation and propose an alternative solution. In particular, we develop a general-purpose primal heuristic to handle MINLPs with nonlinearities in continuous variables.

2 - Computing Clusters in Networks Using Continuous Cubic Formulations

Ashwini Ravindran, Texas A&M University, College Station, TX, United States, Sergiy Butenko

We study alternative continuous cubic formulations for a class of fundamental cluster-detection problems in networks, including the maximum clique, s-defective clique, s-plex, and the independent union of cliques problems. These combinatorial problems, central to graph theory and network analysis, have wide-ranging applications in social network analysis, bioinformatics, and communication systems. The considered formulations capture the structural properties of these discrete problems while enabling their treatment within a continuous optimization framework. Each problem is modeled as the maximization of a degree-three multi-linear polynomial objective over a polyhedral feasible set defined by either the standard simplex or unit hypercube.

These continuous models provide new perspectives for theoretical analysis and algorithmic development. In particular, we derive analytical bounds that offer insight into the quality of feasible solutions and can be used to guide solution refinement. Additionally, we leverage the structure of the cubic formulations to design local and global solution methods based on existing nonlinear optimization solvers.

To assess the practical utility of our approach, we conduct a comprehensive computational study on standard benchmark instances. We compare the performance of multiple solvers in terms of solution quality, computational efficiency, and scalability. Our numerical results demonstrate the effectiveness of the proposed formulations in yielding high-quality solutions and suggest that continuous polynomial models can serve as a powerful alternative to traditional discrete or integer programming techniques for cluster detection in complex networks.

3 - New Methods for Computing the Value Function of Mixed Integer Programs

Siva Ramani, Rice University, Houston, TX, United States, Tu Nguyen, Andrew Schaefer

In mixed integer programming (MIP), the MIP value function (MIPVF) is the optimal value of MIP parametrized by its right-hand side. MIPVF has been leveraged in various algorithmic approaches to solve stochastic MIPs and special cases of bilevel optimization problems and inverse MIPs. We develop a new procedure that exploits the structure of MIPVF in conjunction with results from polyhedral theory and MIP duality to efficiently compute the MIPVF. Our algorithm is scalable and we demonstrate how to leverage our procedure along with other decomposition methods to solve a wide array of problems such as large-scale stochastic MIPs, inverse MIPs, and bilevel programs.

4 - Bilevel Risk-Averse Routing and Resource Allocation for Urban Air Mobility

Haoyun Deng, Georgia Institute of Technology, Atlanta, GA, United States, Weijun Xie

Urban air mobility (UAM) provides a promising solution for alleviating ground traffic congestion and achieving zero direct aviation emissions. The successful and reliable implementation of UAM requires careful coordination between regulatory authorities, such as the Federal Aviation Administration (FAA), and airline operators, particularly in uncertain conditions of passenger demand and airspace capacity. In this paper, we introduce a bilevel risk-averse optimization framework that integrates airline routing decisions with FAA capacity management and delay assignment at critical airspace bottlenecks. To address the computational complexity of this integrated routing and resource allocation problem, we propose and evaluate both heuristic and exact optimization methods. Numerical experiments demonstrate the effectiveness of the proposed approaches and provide valuable insights for stakeholders seeking efficient, equitable, and robust UAM operations in uncertain environments.

MA35

Building B Level 2 B209

Fair Resource Allocation in Non-profit Operations

Invited Session

Public Sector OR

Chair: Haiqing Gao, Northwestern University, Evanston, United States

Co-Chair: Sean Sinclair, Northwestern University, Evanston, IL, United States

1 - Conservative OR Risk-Taking? Objective Selection for Fair Allocation and Routing in Nonprofits**Haiqing Gao, Northwestern University, Evanston, IL, United States, Seyed Iravani, Dawson Ren, Sean Sinclair**

We investigate an online fair sequential allocation problem in which a central planner distributes a divisible resource across multiple locations under demand uncertainty. Motivated by real-world applications, we integrate routing decisions into the planner's problem and formulate dynamic programs (DPs) that jointly optimize allocation and visitation order for two max-min fairness objectives: $\{ \text{ex-post} \}$ and $\{ \text{forward} \}$. Leveraging these formulations, we derive structural insights into the optimal policy. Our main contributions are twofold: First, we show that for any fixed visitation order, the optimal allocation policy follows a $\{ \text{threshold structure} \}$ governed by an $\{ \text{equating property} \}$: locations with sufficiently low demand are fully satisfied, while the remainder receive proportional allocations. Second, we characterize the optimal routing policy and prove that, under mild conditions, it is optimal to visit locations in decreasing order of their demand's coefficient of variation (CV). We complement our theoretical results with a numerical study using real-world data from a food pantry network. Our experiments compare various fairness objectives across multiple fairness metrics, highlighting how each approach navigates the equity-efficiency trade-off. These results offer practical insights into the policy implications of adopting different fairness criteria in dynamic resource allocation.

2 - Coordinator or Active Participant? the Food Banks Role in Ad-Hoc Food Rescue Operations**Irem Sengul Orgut, Loyola University Chicago, Chicago, IL, United States, Shabnam Salehi, Lauren Davis, Emmett Lodree**

In the United States, food banks distribute donated food to individuals and families in need through networks of partner agencies, aiming to ensure fair distribution, minimize waste, and reduce costs. While most donations are received in bulk through scheduled agreements with major donors, food rescue efforts redirect surplus food from smaller-volume donors through ad-hoc, opportunistic donations. These efforts differ from traditional operations in both timing and logistics. This study examines how food banks can best support food rescue activities while balancing geographic equity, effectiveness, and efficiency. We focus on two operational models: a centralized approach, in which the food bank both coordinates and directly participates in rescue efforts, and a decentralized model, where the food bank acts solely as a coordinator and partner agencies handle collection and distribution. Through an extensive numerical study motivated by the real-world operations of our food bank partner, we find that conventional supply chain wisdom favoring centralization does not always apply. When agencies have sufficient storage and rescue capacity, decentralized coordination is more advantageous. In contrast, when agency capacity is low, food bank participation (centralization) leads to better outcomes. We also evaluate food waste impacts and find that centralized models benefit centrally located urban areas, while decentralized systems show no advantage based on proximity to the food bank.

3 - Online Fair Division: Towards Ex-Post Constant Mms Guarantees**Parnian Shahkar, University of California, Irvine, Irvine, CA, United States, Pooja Kulkarni, Ruta Mehta**

We investigate the problem of fairly allocating m indivisible items among n sequentially arriving agents with additive valuations, under the sought-after fairness notion of maximin share (MMS). We first observe a strong impossibility: without appropriate knowledge about the valuation functions of the incoming agents, no online algorithm can ensure any non-trivial MMS approximation, even when there are only two agents. Motivated by this impossibility, we introduce **OnlineKTypeFD** (online k -type fair division), a model that balances theoretical tractability with real-world applicability. In this model, each arriving agent belongs to one of k types, with all agents of a given type sharing the same known valuation function. We do not constrain k to be a constant. Upon arrival, an agent reveals her type, receives an irrevocable allocation, and departs. We study the ex-post MMS guarantees of online algorithms under two arrival models:

1- *Adversarial arrivals*: In this model, an adversary determines the type of each arriving agent. We design a $1/k$ -MMS competitive algorithm and complement it with a lower bound, ruling out any $\Omega(1/\sqrt{k})$ -MMS-competitive algorithm, even for binary valuations.

2- *Stochastic arrivals*: In this model, the type of each arriving agent is independently drawn from an underlying, possibly unknown distribution. Unlike the adversarial setting where the dependence on k is unavoidable, we surprisingly show that in the stochastic setting, an asymptotic, arbitrarily close-to- $1/2$ -MMS competitive guarantee is achievable under mild distributional assumptions.

Our results extend naturally to a learning-augmented framework as the competitive ratios of our algorithms degrade gracefully with multiplicative prediction errors.

4 - Forward-Backward Contention Resolution Schemes For Fair Rationing**Will Ma, Columbia Business School, New York, NY, United States, Calum MacRury, Clifford Stein**

We use contention resolution schemes (CRS) to derive algorithms for the fair rationing of a single resource when agents have stochastic demands. We aim to provide ex-ante guarantees on the level of service provided to each agent, who may measure service in different ways (Type-I, II, or III), calling for CRS under different feasibility constraints (rank-1 matroid or knapsack). We are particularly interested in two-order CRS where the agents are equally likely to arrive in a known forward order or its reverse, which is motivated by online rationing at food banks.

In particular, we derive a two-order CRS for rank-1 matroids with guarantee $1/(1+e^{-1/2}) \approx 0.622$, which we prove is tight. This improves upon the $1/2$ guarantee that is best-possible under a single order (Alaei, SIAM J. Comput. 2014), while achieving separation with the $1-1/e \approx 0.632$ guarantee that is possible for random-order CRS (Lee and Singla, ESA 2018). Because CRS guarantees imply prophet inequalities, this also beats the two-order prophet inequality with ratio $(\sqrt{5}-1)/2 \approx 0.618$ from (Arsenis, SODA 2021), which was tight for single-threshold policies. Rank-1 matroids suffice to provide guarantees under Type-II or III service, but Type-I service requires knapsack. Accordingly, we derive a two-order CRS for knapsack with guarantee $1/3$, improving upon the $1/(3+e^{-2}) \approx 0.319$ guarantee that is best-possible under a single order (Jiang et al., SODA 2022). To our knowledge, $1/3$ provides the best-known guarantee for knapsack CRS even in the offline setting. Finally, we provide an upper bound of $1/(2+e^{-1}) \approx 0.422$ for two-order knapsack CRS.

MA36

Building B Level 2 B210

Stochastic and Noisy Algorithms for Continuous Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Frank E. Curtis, Lehigh University, Bethlehem, PA, United States

1 - Negative Curvature Methods with High Probability Complexity Bounds for Stochastic Nonconvex Optimization**Wanping Dong, University of Michigan, Ann Arbor, MI, United States, Albert Berahas, Raghu Bollapragada**

We consider two-step negative curvature methods for solving nonlinear nonconvex unconstrained optimization problems, where function, gradient, and Hessian evaluations are provided by oracles. We establish second-order convergence results under two types of oracle models. First, we analyze two-step methods with constant step sizes under deterministic inexact oracles, deriving deterministic second-order guarantees and iteration complexity bounds under reasonable assumptions. Building on this, we extend the analysis to probabilistic oracles, where the function, gradient, and Hessian estimates satisfy accuracy conditions with high probability. The algorithm we propose incorporates an early stopping mechanism to ensure sufficient progress at each step and a step search strategy based on Armijo-type conditions to dynamically

adjust step sizes for both steps. These strategies enable high-probability second-order convergence guarantees and iteration complexity bounds. Numerical results on machine learning problems demonstrate the efficiency of the proposed method in practice.

2 - A Sequential Cubic Optimization Method with Second-Order Complexity Guarantees for Equality Constrained Optimization

Nikolaos Dimou, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

We develop a new method for equality constrained optimization based on a sequential cubic optimization framework. Each iteration utilizes a step decomposition based on the Jacobian of the constraints into a normal and a tangential component, the latter of which is found by solving a subproblem involving cubic regularization. The method incorporates second order correction steps as necessary to ensure global convergence to second order stationary points, as well as local quadratic convergence. In addition, we show that the algorithm is the first to obtain worst-case complexity guarantees on the order of $\epsilon^{-3/2}$ for the gradient of the Lagrangian, ϵ^{-3} in terms of second order stationarity, and ϵ^{-1} in terms of the constraint violation. These complexity guarantees are the best known results of any method for this class of problems.

3 - Progressive Sampling for Equality Constrained Optimization with Expectation Constraints

Lingjun Guo, Lehigh University, Bethlehem, PA, United States, Frank Curtis, Daniel Robinson

We consider the problem to minimize a smooth objective function subject to equality constraints. In particular, we are motivated by the setting of constrained supervised machine learning where the aim is to minimize an empirical loss function subject to constraints defined by an expectation or large finite sum. In this context, utilizing exact constraint gradient information requires a pass over all samples in the dataset, which can be prohibitively expensive when the number of samples is large. Thus, we propose a progressive sampling strategy that initially solves a subproblem whose constraints only use a small subset of the full samples, then progressively increases the size of the subset while solving each subproblem approximately until the solution reaches the desired accuracy. Numerical simulations demonstrate that our method can outperform an approach that solves the original problem directly when performance is measured by the number of sample constraint gradients requires to obtain a solution of comparable quality.

4 - Adaptive Stochastic Gradient Tracking Methods for Decentralized Optimization

Wei Wang, The University of Texas at Austin, Austin, TX, United States, Albert Berahas, Raghu Bollapragada

In decentralized optimization, multiple nodes in a network collectively solve an optimization problem over a global objective function. Gradient tracking methods have emerged as a popular approach in this subfield, as they offer both strong theoretical guarantees and robust empirical performance. In these methods, the complexity of the two operational steps—computing local gradients and communicating between nodes—must be balanced for various large-scale stochastic optimization applications. In this work, we adopt an adaptive sampling method using a well-known norm test within a stochastic version of a flexible, unified gradient tracking framework. The proposed framework unifies existing gradient tracking methods while dynamically controlling sampling selection for local gradient approximation at each node and employing a randomized communication scheme. We establish convergence guarantees in expectation and analyze the complexity of both computation and communication. Numerical experiments on logistic regression and quadratic problems illustrate the effectiveness of the proposed framework.

5 - Active-Set Identification in Noisy and Stochastic Algorithms

Lara Zebiane, Lehigh University, Bethlehem, PA, United States, Frank E. Curtis, Daniel Robinson

We will discuss theoretical guarantees pertaining to constrained optimization problems. In particular, we will focus on noisy and stochastic methods and their ability to identify optimal active-constraint sets. A major challenge in this context is that, since the function and derivative values have error, standard techniques might fail to identify the correct active-constraint set even at a local minimizer. We will show that, even in the presence of errors, estimates can be computed that will identify the correct active-constraint set when the errors are accounted for appropriately. We also provide some numerical demonstration of our theoretical guarantees.

MA37

Building B Level 2 B211

Data-Driven and Distributionally-Robust Optimization under Distribution Shifts

Invited Session

OPT: Optimization Under Uncertainty

Chair: Dominic Keehan, University of Auckland, Auckland, New Zealand

1 - Incorporating Unknown Trust into Decision Recommendation with Multi-Sourced Reference Information

Yanru Guo, University of Michigan, Ann Arbor, MI, United States, Ruiwei Jiang, Siqian Shen

Navigating the complexities of decision-making processes relies heavily on the ability to integrate information from multiple sources, especially in environments where uncertainty exists. In this study, we investigate a stochastic optimization problem where the probability distributions of uncertain parameters are estimated using reference information from various sources. We conduct non-parametric data fusion based on trust over different sources to construct the Wasserstein ambiguity set and formulate a distributionally robust optimization (DRO) model to minimize potential worst-case losses. In addition, we design a dynamic trust update framework and propose multiple update rules to simulate decision maker's trust variation. We demonstrate the effectiveness of our approaches using instances of preventive care allocation.

2 - Don't Look Back in Anger: Wasserstein Distributionally Robust Optimization with Nonstationary Data

Dominic Keehan, University of Auckland, Auckland, New Zealand

We study data-driven decision-making problems in which historical data is sampled from a sequence of distributions that change over time. Every consecutive distribution is shifted by at most a certain Wasserstein distance. To address this nonstationarity, we propose a distributionally robust optimization with a Wasserstein-ball based ambiguity set. At the center of this ball is a weighted empirical distribution on the historical data. By deriving a concentration inequality for weighted empiricals which explicitly incorporates the possible distribution shift, we optimize the datum weightings, and thereby maximize the probability that our ambiguity ball contains the next distribution in the sequence. When there is no distribution shift, our approach naturally recovers standard Wasserstein distributionally robust optimization. Moreover, it reveals a relationship between the order of the Wasserstein ambiguity ball and the decay profile of the optimal datum weightings. Comparison of the resulting weightings shows that simple weighting approaches like windowing and exponential smoothing are near optimal. Numerical experiments demonstrate the effectiveness of our approach.

3 - Nonstationary Distribution Estimation via Wasserstein Probability Flows

Edward Anderson, Imperial College London, Cambridge, United Kingdom, Dominic Keehan

We consider a situation in which historical observations are drawn from an underlying probability distribution which changes over time. We model changes in the distribution nonparametrically, and suppose that changes are likely to involve a small Wasserstein distance between two consecutive distributions. This leads naturally to a formulation in which we estimate the underlying series of distributions by maximizing the log likelihood of the observations with a penalty applied to the sum of the Wasserstein distances between consecutive distributions. We call this the Wasserstein Probability Flow problem, and show how it reduces to a simple network-flow problem. This can be easily solved, and we derive some properties of its optimal solutions. A range of numerical experiments in different settings show that the Wasserstein Probability Flow method for estimating distributions can be effective in practice.

4 - Ambiguity Updating Robust Architecture: Updating Ambiguity Sets with Revealed Uncertainty in Multi-Period Dro

Gar Goei Loke, Durham University Business School, Durham, United Kingdom, Wenxin Liu, Yu Zhang, Aakil Caunhye

We study the context of multi-period distributionally robust optimization, where uncertainties are revealed sequentially and the decision-maker wishes to re-solve the problem-to-go, utilizing this information to shrink the ambiguity set of distributions-to-go, while remaining consistent with the original ambiguity set. We propose a model, termed Ambiguity Updating Robust Architecture (AURA), that defines the ambiguity sets for the problems-to-go. Here, AURA does not require uncertainties to be

independent across time, or explicit characterization of the conditional distribution of the uncertainties on those at prior times. We show that the robust counterpart of AURA can be tractably computed. Under linear assumptions, AURA can be reformulated as a linear optimization model with number of decision variables and constraints that grows quadratically in time.

5 - Towards Optimal Offline Reinforcement Learning

Mengmeng Li, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland, Daniel Kuhn, Tobias Sutter

We study offline reinforcement learning problems with a long-run average reward objective. The state-action pairs generated by any fixed behavioral policy thus follow a Markov chain, and the empirical state-action-next-state distribution satisfies a large deviations principle. We use the rate function of this large deviations principle to construct an uncertainty set for the unknown true state-action-next-state distribution. We also construct a distribution shift transformation that maps any distribution in this uncertainty set to a state-action-next-state distribution of the Markov chain generated by a fixed evaluation policy, which may differ from the unknown behavioral policy. We prove that the worst-case average reward of the evaluation policy with respect to all distributions in the shifted uncertainty set provides, in a rigorous statistical sense, the least conservative estimator for the average reward under the unknown true distribution. This guarantee is available even if one has only access to one single trajectory of serially correlated state-action pairs. The emerging robust optimization problem can be viewed as a robust Markov decision process with a non-rectangular uncertainty set. We adapt an efficient policy gradient algorithm to solve this problem. Numerical experiments show that our methods compare favorably against state-of-the-art methods.

MA38

Building B Level 2 B212

Network Optimization in Power System Operations and Planning

Invited Session

OPT: Network Optimization

Chair: Behnam Jabbari Marand, North Carolina State University, Raleigh, NC, 27606, United States

1 - Path-Based Valid Inequalities for Voltage Angle Differences in Optimal Transmission Switching

Behnam Jabbari Marand, North Carolina State University, Raleigh, NC, United States, Adolfo Escobedo

The switching of transmission lines can significantly enhance the economic and operational efficiency of power systems. The Direct-Current Optimal Transmission Switching (DC-OTS) problem which minimizes power generation costs through reconfiguration of the transmission network under the DC power flow model, provides a formal optimization framework for making these strategic decisions. DC-OTS is typically formulated as a mixed-integer linear program, incorporating disjunctive constraints to enforce Kirchhoff's Voltage Law (KVL) via big-\$M\$ parameters. In angle-based formulations, these parameters represent upper bounds on voltage angle differences across transmission lines; yet in practice, overly conservative (and arbitrary) bounds tend to be used. The belief is that tightening these values requires the solution of the longest path problem, which is computationally intractable. In this work, we present a more efficient and theoretically grounded alternative. To that end, we derive an extended formulation of the convex hull of a relaxation of the DC-OTS feasible region. From this formulation, we obtain strong path-based valid inequalities that tighten big-\$M\$ parameters and compare them against the cycle-based inequalities proposed by Kocuk et al. 2016. Computational experiments on standard benchmark instances show that incorporating these inequalities within a cutting-plane framework substantially improves solver performance, demonstrating their practical effectiveness for solving large-scale DC-OTS problems.

2 - Mixed Integer Programming Formulations for Optimal Power Flow

Gonzalo Gomez Abejon, Rice University, Houston, TX, United States, Illya Hicks

Optimal Power Flow in AC electrical grids is a nonconvex problem for which it can be notably difficult to obtain tight dual bounds and certify global optimality. Lower bounds are typically found by solving convex relaxations, which can only approximate the original constraints to a certain extent.

Instead, we explore mixed integer programming relaxations that approximate the nonconvex constraints by a union of n polytopes and can produce an arbitrarily small duality gap if solved exactly. We propose MILP formulations that only require $O(\log(n))$ total variables and constraints to approximate certain nonlinear constraints, an exponential reduction over previous formulations, and prove that this bound is asymptotically optimal, that is, that we can only reduce the formulation size further by a constant factor. Finally, we present some computational experiments where the performance of MIP relaxations is compared with standard convex relaxations for AC Optimal Power Flow.

3 - High-Spatial Resolution Transmission and Storage Expansion Planning for High-Renewable Power Systems: a Case Study

Kevin Wu, Georgia Institute of Technology, Atlanta, GA, United States, Rabab Haider, Pascal Van Hentenryck

The U.S. electric grid faces increasing challenges due to growing renewable energy integration and rising electricity demand. To ensure reliable and cost-effective power delivery under these conditions, long-term Transmission Expansion Planning (TEP) and energy storage deployment must be coordinated. This paper presents a scalable, high-resolution framework for joint TEP and storage planning, addressing tractability limitations that have constrained prior work to smaller or more aggregated systems.

A multiperiod power transfer distribution factor (PTDF)-based DC optimal power flow (DCOPF) formulation is developed, enabling direct modeling of storage impacts on line flows over a 24-hour horizon. To improve computational performance, flow constraints are added lazily in batches sorted by violation severity. For broader scalability across multiple representative days, Benders decomposition is employed with several convergence accelerators, including in-out stabilization and core-point perturbations.

This approach is demonstrated on a 2,000-bus synthetic Texas system under 2050 decarbonization scenarios. The results highlight the critical role of co-optimized storage and transmission in reducing congestion, improving reliability, and lowering system costs. This work offers a tractable pathway for high-resolution long-term planning in future renewable-heavy grids.

4 - Power System State Estimation by Phase Synchronization and Eigenvectors

Iven Guzel, University of Illinois Urbana-Champaign, Urbana, IL, United States, Richard Zhang

To estimate accurate voltage phasors from inaccurate voltage magnitude and complex power measurements, the standard approach is to iteratively refine a good initial guess using the Gauss-Newton method. But the nonconvexity of the estimation makes the Gauss-Newton method sensitive to its initial guess, so human intervention is needed to detect convergence to plausible but ultimately spurious estimates. This paper makes a novel connection between the angle estimation subproblem and phase synchronization to yield two key benefits: (1) an exceptionally high quality initial guess over the angles, known as a spectral initialization; (2) a correctness guarantee for the estimated angles, known as a global optimality certificate. These are formulated as sparse eigenvalue-eigenvector problems, which we efficiently compute in time comparable to a few Gauss-Newton iterations. Our experiments on the complete set of Polish, PEGASE, and RTE models show, where voltage magnitudes are already reasonably accurate, that spectral initialization provides an almost-perfect single-shot estimation of angles from moderately noisy bus power measurements (i.e. pairs of PQ measurements), whose correctness becomes guaranteed after a single Gauss-Newton iteration. For less accurate voltage magnitudes, the performance of the method degrades gracefully; even with moderate voltage magnitude errors, the estimated voltage angles remain surprisingly accurate.

5 - Stochastic Nodal Power System Capacity Expansion Planning with Siting of Large Loads

Tomas Valencia Zuluaga, Lawrence Livermore National Laboratory, San Francisco, CA, United States, Simon Pang, Jean-Paul Watson

Traditional power grid expansion planning studies assume that load growth is an exogenous parameter of the optimization model. The proliferation of datacenters supporting the so-called AI revolution, as well as industrial facilities built for direct air capture of CO₂ are challenging this view. These loads can be sited with some flexibility and are large enough to have an impact on the electricity network, even at the transmission level. In this work, we incorporate large-scale loads with flexible siting as part of the decision variables in a nodal capacity expansion planning model and implement a decomposition methodology to solve this optimization problem exploiting parallel computing. We show through some numerical testcases on realistic systems based on California and South Carolina the impact of proactive load siting on the need for transmission investments with different power flow formulations to highlight the importance of incorporating them into software planning tools.

MA39

Building B Level 2 B213

GPU-Based Mathematical Optimization

Invited Session

OPT: Linear and Conic Optimization

Chair: Haihao Lu, MIT, Chicago, IL, 02139, United States

Co-Chair: Jinwen Yang, University of Chicago, Chicago, IL, United States

1 - Gpu-Accelerated Linear Programming and Beyond**Jinwen Yang, University of Chicago, Chicago, IL, United States**, Haihao Lu

The rapid progress in GPU computing has revolutionized many fields, yet its potential in mathematical programming, such as linear programming (LP), has only recently begun to be realized. This talk aims to provide an overview of recent advancements in GPU-based first-order methods for LP, with a particular focus on the design and development of cuPDLP. The extensions to GPU-based optimization beyond LP, including quadratic, semidefinite and conic programming, will also be discussed.

2 - Mpx: Mathematical Programming in Jax**Zedong Peng, MIT, Cambridge, MA, United States**, Haihao Lu, Jinwen Yang

We introduce MPAX (Mathematical Programming in JAX), a versatile and efficient toolbox for integrating mathematical programming into machine learning workflows. MPAX implemented the state-of-the-art first-order methods, restarted average primal-dual hybrid gradient and reflected restarted Halpern primal-dual hybrid gradient, to solve linear programming and quadratic programming problems in JAX. It provides native support for hardware accelerations along with features like batch solving, auto-differentiation, and device parallelism. Extensive numerical experiments demonstrate the advantages of MPAX over existing solvers. The solver is available at <https://github.com/MIT-Lu-Lab/MPAX>.

3 - Large-Scale Semidefinite Programming with Gpu-Accelerated Admm**Heng Yang, Harvard University, Boston, MA, United States**, Shucheng Kang, Xin Jiang, Antoine Groudiev

Large-scale semidefinite programs (SDPs) are increasingly important due to the growing use of semidefinite relaxations—particularly the Moment-SOS hierarchy—in engineering and applied sciences.

In this talk, we show that the Alternating Direction Method of Multipliers (ADMM) can serve as a powerful and broadly applicable first-order algorithm for solving large-scale SDPs.

First, we present a new theoretical result that challenges the common perception that ADMM is limited to medium-accuracy solutions. Specifically, we prove that if ADMM converges to a primal-dual solution pair satisfying strict complementarity, it achieves local linear convergence. Extensive numerical experiments on a wide variety of SDP instances support our theory and demonstrate that ADMM can reach high-accuracy solutions—even in the absence of nondegeneracy conditions.

Second, we present cuADMM, a generic GPU-accelerated ADMM-based solver for large-scale SDPs. We demonstrate its superior performance compared to existing SDP solvers, highlighting its potential for scaling semidefinite optimization to previously intractable problem sizes.

4 - Solving Large-Scale, Sparse, Constrained Nonlinear Optimization Problems on Gpus**Sungho Shin, Massachusetts Institute of Technology, Cambridge, MA, United States**, François Pacaud, Alexis Montoisson

In this talk, we delve into recent advancements in solving large-scale sparse nonlinear optimization problems on GPUs. This new approach combines three key elements: (1) the SIMD abstraction of nonlinear optimization problems, (2) the condensed-space interior point method approach, and (3) NVIDIA's recently released sparse linear solver, CUDSS. By combining these components, we provide a comprehensive nonlinear optimization solution capabilities. We present numerical experiment results showcasing over 10x speed-ups when applying this method to large-scale AC optimal power flow problems. We will conclude the discussion by examining the potential implications of these findings for the future of power system optimization and decision-making.

5 - Pdc: a Primal-Dual Large-Scale Conic Programming Solver with Gpu Enhancements**Zikai Xiong, Operations Research Center, MIT, Cambridge, MA, United States**, Zhenwei Lin, Dongdong Ge, Yinyu Ye

In this talk, we introduce the Primal-Dual Conic Programming Solver (PDCS), a large-scale conic programming solver with GPU enhancements. Problems that PDCS currently supports include linear programs, second-order cone programs, convex quadratic programs, and exponential cone programs. PDCS achieves scalability to large-scale problems by leveraging sparse matrix-vector multiplication as its core computational operation, which is both memory-efficient and well-suited for GPU acceleration. The solver is based on the restarted primal-dual hybrid gradient method but further incorporates several enhancements, including adaptive reflected Halpern restarts, adaptive step-size selection, adaptive weight adjustment, and diagonal rescaling. Additionally, PDCS employs a bijection-based method to compute projections onto rescaled cones. Furthermore, cuPDCS is a GPU implementation of PDCS and it implements customized computational schemes that utilize different levels of GPU architecture to handle cones of different types and sizes. Numerical experiments demonstrate that cuPDCS is generally more efficient than state-of-the-art commercial solvers and other first-order methods on large-scale conic program applications, including Fisher market equilibrium problems, Lasso regression, and multi-period portfolio optimization. Furthermore, cuPDCS also exhibits better scalability, efficiency, and robustness compared to other first-order methods on the conic program benchmark dataset CBLIB. These advantages are more pronounced in large-scale, lower-accuracy settings.

MA40

Building B Level 2 B214

Trust and Optimization in Public-Interest Systems: AI, Mobility, and Energy

Invited Session

Service Science

Chair: Varsha Shukla, Dartmouth College, Hanover, NH, 03755, United States

1 - Accelerating Airport Decarbonization Through Multi-Stage Stochastic Optimization and Vehicle-to-Grid Integration and Renewables**Benjamin Amoh, Dartmouth College, Lebanon, NH, United States**

Airports are energy-intensive infrastructures with growing pressure to decarbonize amidst rising electrification demands and sustainability mandates. This presentation introduces a comprehensive decision-support framework for optimizing airport energy transitions using a multi-stage stochastic optimization model. Our approach simultaneously addresses investment planning in renewable energy (solar and wind), battery energy storage, and electric ground support equipment (e-GSE), while accounting for deep uncertainties in weather and demand over a multi-decade horizon.

To tackle the complexity of high-dimensional temporal uncertainty, we apply k-means clustering on 49 years of hourly solar irradiance and wind speed data to construct 15 representative weather scenarios, weighted monthly using a Wasserstein distance approach. These scenarios drive our stochastic model, capturing seasonal and long-

term climate variability. We further incorporate vehicle-to-grid (V2G) strategies by modeling the battery capacities of airport-owned EV fleets—including buses, pushback tractors, and baggage loaders—while accounting for battery degradation dynamics. This enables airports to leverage existing mobile storage assets to reduce peak demand charges and support grid stability. Our model is applied to several major U.S. airports with distinct energy profiles, revealing optimal investment pathways that balance emissions reduction, cost-effectiveness, and operational reliability. Results show that integrating renewables, V2G, and robust planning techniques can yield up to 35% cost savings and 60% emission reductions compared to static strategies. This work demonstrates a scalable, data-driven framework that aligns infrastructure decarbonization with resilience and efficiency—offering actionable insights for airport authorities, energy planners, and policymakers.

2 - Optimizing Shared Micromobility Service Network Design

Vikrant Vaze, Dartmouth College, Hanover, NH, United States, Pushendra Singh

Shared micromobility systems (SMS), including bike and e-scooter sharing, support the shift toward sustainable urban mobility. However, their financial viability hinges on balancing capital investment and service reliability. Oversized fleets and stations waste resources, while shortages reduce customer satisfaction. This paper presents a mathematical and computational framework to optimize SMS design by aligning capacity decisions with customer demand. We formulate a mixed-integer non-convex optimization model to determine station locations, capacities, and micromobility and rebalancing vehicle fleet sizes. Our model integrates customer preferences via a discrete choice framework and captures key operational and rebalancing dynamics.

We develop an original spatial decomposition heuristic. It separates the overall model into station-location-specific models and couples operations at individual stations through iterative bilevel updates. Local updates are computationally less expensive, occur at a higher frequency, and improve station-specific solutions, whereas the global updates are computationally more expensive, occur at a lower frequency, and reset the errors accumulated over multiple local updates to drive the solution toward optimality.

Extensive experiments using real-world data show that our approach delivers near-optimal solutions, while reducing computational times by several orders of magnitude. This scalability enables system-wide optimization, serving up to 14% more demand, and generating 82% higher profitability, and 17% lower rebalancing costs than current practice. Ultimately, our comprehensive optimization formulation combined with our efficient solution approach provides bottomline improvement annually worth 8-40 million dollars to the system operator, enhances system accessibility to the customers, and reduces environmental footprint, resulting in a more profitable, accessible and sustainable micromobility system.

3 - Multi-Objective, Large-Scale, Stochastic Optimization for Green Energy Infrastructure to Service Conflict-Affected Regions of Africa

Lilly Yang, Dartmouth College, Hanover, NH, United States, Siqi Ke, Vikrant Vaze

Fragile states struggle with chronically low electrification rates, acute vulnerability to the effects of climate change, and heavy dependence on imported diesel that frequently which often leads to violence and conflict related to fuel transport. United Nations (UN) peacekeeping force compounds often correspond to a disproportionate share of the national electricity consumption, while adjacent communities remain largely underserved—highlighting the need to leapfrog the traditional fossil-fuel based grid development and adopt decentralized renewable solutions. This study presents a two-stage stochastic, multi-objective optimization framework for designing and operating decentralized hybrid renewable-energy systems (D-HRES) in fragile states. The model jointly optimizes upfront investment, diesel consumption, community coverage, and supply reliability. Uncertainty in renewable availability is captured through a scenario set reduced from 20 years of weather data via exact integer programming. First stage of the stochastic optimization framework determines the technology mix and installed capacities, whereas the second-stage decisions optimize hourly dispatch for each realized scenario. This multi-objective, large-scale, stochastic optimization problem is solved with an ϵ -constraints solution approach, and results are benchmarked against true optima as well as against a state-of-the-art multi-objective metaheuristic approach (called NSGA-II). Case studies across sub-Saharan Africa, supported by site-specific data from field experts, demonstrate the framework's transferability and yield actionable trade-off insights for planners and humanitarian agencies.

4 - Policy-Aligned Generative AI: a Practical Playbook for Trust in Public-Interest Systems

Varsha Shukla, Dartmouth College, Hanover, NH, United States

Public agencies and regulated enterprises are adopting generative AI, but many deployments stall on questions of safety, bias, privacy, and accountability. This talk offers a practical playbook for making generative AI policy-aligned and trustworthy in public-interest domains such as health, mobility, and energy. I translate high-level principles (fairness, transparency, human oversight, data minimization) into concrete engineering and product practices: role-based access, PHI-safe data paths, model and data provenance, evaluation plans tied to use-case risk, red-team protocols, audit logs, and clear rollback procedures.

Using a healthcare analytics copilot as a case example, I show how to combine a semantic data layer with guardrails (schema-aware text-to-SQL, codebook versioning, pre-execution checks) and with governance assets (model cards, risk registers, incident response runbooks). I then outline how the same controls generalize to mobility (service design under equity constraints) and energy (planning under uncertainty), where decisions carry real-world impact.

Attendees will leave with a checklist that links policy goals to measurable controls, what to track before launch (intended use, data lineage, evaluation criteria), during operation (drift, access, cost/safety thresholds), and after (auditability and remediation). The emphasis is on simple, repeatable mechanisms that raise trust while keeping teams productive and within budget.

MA41

Building B Level 3 B301

OR/MS applications to improve women's health

Invited Session

Health Applications Society

Chair: Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States

1 - Informing Breast Cancer Control Policy Through Mathematical Modeling: the Cancer Intervention and Modeling Surveillance Modeling (Cisnet) Experience

Oguzhan Alagoz, University of Wisconsin-Madison, Madison, WI, United States

Breast cancer remains the most common non-skin cancer and the second leading cause of cancer-related death among women in the United States. While mammography continues to be the most effective diagnostic tool for breast cancer detection, it involves several risks, including a relatively high rate of false positives. Therefore, a careful balance of benefits and harms is essential in the design of mammography screening schedules.

This presentation focuses on the pivotal role of mathematical modeling in optimizing breast cancer screening strategies, with a particular emphasis on the University of Wisconsin Breast Cancer Simulation Model (UWBCS), a member of the National Cancer Institute's Cancer Intervention and Surveillance Modeling Network (CISNET). The UWBCS integrates extensive data sources, including SEER, BCSC, NHIS, and the clinical literature, to simulate real-world screening behaviors, tumor growth dynamics, detection mechanisms, and treatment dissemination. The model successfully replicated age-, subtype-, and stage-specific incidence, survival, and mortality patterns reported in SEER and externally validated against Wisconsin Cancer Reporting System and the UK AGE Screening Trial.

Real-world policy decisions have been directly informed by the UWBCS. Most significantly, the model was employed by the U.S. Preventive Services Task Force, alongside other CISNET models, to guide the national breast cancer screening recommendations in 2009, 2016, and 2024. In addition, it impacted the 2024 screening

guidelines by the National Comprehensive Cancer Network for women with high-risk genetic variants. The model was recently adopted by the Canadian Partnership Against Cancer.

2 - Modeling the Natural History of Ovarian Cancer

Iakovos Toumazis, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, Sayaka Ishizawa, Larissa A Meyer

Background: Ovarian cancer is the second leading cause of death among all gynecologic cancers. Yet, no effective prevention program currently exists. Well-validated natural history models (NHMs) provide insights on the natural progression of diseases and can be used to assess the effectiveness of potential interventions in silico. Currently, no histology-specific NHM exist for ovarian cancer.

Methods: We developed histology-specific NHMs for the seven most common ovarian cancer subtypes using discrete-time Markov models. We calibrated histology-specific model parameters using Adaptive Simulated Annealing such that model outcomes reliably reproduced the observations from the Surveillance, Epidemiology, and End Results (SEER) program. We assessed goodness-of-fit using the weighted sum of root mean squared error (RMSE). We validated the NHMs on data from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) and the United Kingdom Collaborative Trial on Ovarian Cancer Screening (UKCTOCS). Moreover, the stage distribution at diagnosis was validated on PLCO data. Differences between observations and model estimates were evaluated using the Pearson's chi-squared test.

Results: The calibrated NHM showed satisfactory goodness of fit to the SEER data, within 2% of RMSE. In model validation, none of our model estimates showed significant difference on UKCTOCS (p-value on incidence=0.399, death=0.377) or PLCO (p-value on incidence=0.207, death=0.513, stage distribution=0.498) observations.

Conclusion: The developed NHM accurately describes the histology-specific natural progression of ovarian cancer. This model could be used to assess the impact of future and emerging ovarian cancer preventive interventions, thus assisting policy makers developed effective prevention programs.

3 - Political Leadership and Contraceptive Uptake

Dwaipayan Roy, University of Virginia Darden School of Business, Charlottesville, VA, United States, Amir Karimi

In low- and middle-income countries (LMICs), government health facilities are the primary providers of contraceptives, especially for the poorest populations. In India, over 85% of the lowest-income users rely on government health facilities for contraception. However, contraceptive uptake in India remains low—almost 50 million women fall into this category, accounting for 90% of unintended pregnancies. This study examines whether publicly elected female leaders affect contraceptive uptake in India's government health facilities. Using a quasi-experimental design based on close elections in 692 districts over a decade, we find that female legislators increase the uptake of intra-uterine devices (IUDs), a long-acting reversible method. Additional analysis reveals that female legislators boost contraceptive counseling by frontline health workers, expand contraceptive awareness through mass media, and increase postpartum sterilizations. These pathways suggest that female leaders promote reproductive health (RH) in their communities more than their male peers. Our findings highlight female political leadership as a key operational lever to improve RH outcomes. With India's recent legislation to reserve one-third of assembly seats for women, similar reforms in other LMICs could help enhance contraceptive uptake. These insights also contribute to the global health operations literature by identifying female political leadership as a critical factor in improving contraceptive uptake.

4 - Examining Perinatal Regionalization in Practice: a Network Analysis of Maternal Transport in Georgia

Jingyu Li, Georgia Institute of Technology, Atlanta, GA, United States, Lauren Steimle, Stephanie Radke

Perinatal regionalization is a systems-level strategy for coordinating care to ensure pregnant people receive timely care in facilities with risk-appropriate personnel and services. Inter-facility transport of patients to risk-appropriate facilities is a critical component of regionalized systems. As regionalized systems of maternal care are recently emerging, it remains unclear how maternal transport is being operationalized in practice. Using network analysis, we constructed maternal transport network to represent maternal transport routes (network edges) among obstetric facilities (network nodes) as directional network. We characterized patients and regions that were more likely to be involved in transport, and those more likely to be transported across the state's designated perinatal regions. Network metrics were used to discern dynamics and structure of maternal transport in each perinatal region. Finally, we applied Louvain community detection algorithm to cluster obstetric facilities into communities that are observed to transport among each other most frequently. We compared these algorithm-detected communities with state's formal designated perinatal regions. We demonstrated this data-driven approach using maternal transports as indicated on 2017-2022 birth records from the state of Georgia. We identified 8 algorithm-detected communities within each of which transports were coordinated frequently. While these detected communities were largely aligned with the state's designated perinatal regions, geographical proximity between facilities and transporting within health systems tend to trump formal perinatal region designations. Redesigning the state's perinatal region to align with the observed maternal transport networks and to formalize relationships that are naturally occurring in practice could potentially improve access to risk-appropriate maternal care.

MA42

Building B Level 3 B302

Advancing Healthcare Operations through Optimization and Analytics

Invited Session

Health Applications Society

Chair: Yu Ma, University of Wisconsin Madison, Madison, WI, United States

1 - Fairness-Promoting Integer Programming Approaches for Medical Resident Rotation Scheduling

Karmel Shehadeh, University of Southern California, Los Angeles, CA, United States, Shutian Li, Frank Curtis, Beth Hochman

Motivated by our collaboration with a residency program at an academic health system, we propose new integer programming (IP) approaches for the resident-to-rotation assignment problem (RRAP). We first present an IP formulation for the RRAP, which mimics the manual method for generating rotation schedules in practice and can be easily implemented and efficiently solved using off-the-shelf optimization software. However, it can lead to disparities in satisfying vacation requests among residents. To mitigate such disparities, we derive a fairness-promoting counterpart, which finds an optimal rotation schedule that maximizes the number of satisfied vacation requests and minimizes a measure of disparity in satisfying vacation requests. Then, we propose a computationally efficient Pareto Search Algorithm capable of finding the complete set of Pareto optimal solutions to the equity-promoting IP within a time that is suitable for practical implementation. In addition, we present a user-friendly tool implementing the proposed models to automate the generation of the rotation schedule. Finally, we construct diverse RRAP instances based on data from our collaborator and conduct extensive experiments to illustrate the potential practical benefits of our proposed approaches. Our results demonstrate the computational efficiency and implementability of our approaches and underscore their potential to enhance fairness in resident rotation scheduling.

2 - Enhancing Safety Signaling: Integrating Clinical Trials and Post-Marketing Adverse Event Reports

Fernanda Bravo, UCLA Anderson School of Management, Los Angeles, CA, United States, Yunliang Chen, John Silberholz

Adverse drug reactions (ADRs) are harmful side effects that can lead to emergency room visits and fatalities. While regulators like the FDA aim to detect ADRs during clinical trials before a drug reaches the market, these trials often lack the size or duration to catch rare or delayed effects. As a result, post-approval surveillance systems, such as the FDA's FAERS, play a critical role in detecting safety signals—potential indicators of harmful side effects. However, these observational systems suffer from biases due to variations in how and when patients report side effects, potentially leading to both false alarms and overlooked risks.

To improve this process, we propose a Bayesian framework that integrates data from pre-marketing clinical trials with post-marketing observational data. The method estimates and corrects biases in the observational data using patterns observed in more common ADRs, under the assumption that both common and rare ADRs are affected by similar reporting and selection biases. The proposed approach allows for more reliable hypothesis generation about drug safety issues. Through analytical

modeling and empirical testing using real clinical and FAERS data, we show that our method can significantly reduce the likelihood of false positives and missed true ADRs. Specifically, it can lower expected costs by 1-3% rel. to an alternative Bayesian approach that ignores common ADRs. For regulatory agencies, this methodology offers a practical and impactful enhancement to post-market drug safety monitoring, improving the accuracy and speed of ADR detection using already available data.

3 - Provider Payment Models for Transformative Technologies in Healthcare

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States, Elodie Adida

Transformative technologies such as robotic surgery hold promise for improving medical practice and patient outcomes. Despite this potential, their adoption in healthcare is constrained by high costs, the need for customization, and the absence of suitable payment models. This paper examines how provider payment models influence both the adoption of new technologies (downstream) and the quality of these tools (upstream).

We develop a game theoretical framework to analyze the strategic interaction between a developer—who sets the price and determines the tool's quality—and a healthcare provider, who decides on usage based on the reimbursement structure and the expected benefit from the technology. Under the current norm of no reimbursement, providers restrict usage to complex cases, resulting in limited adoption and suboptimal quality. Fee-for-service models encourage broader usage but create perverse incentives: high reimbursement rates prompt developers to reduce quality, while low rates may induce higher quality but fall short of socially optimal outcomes. We propose a hybrid model that blends fee-for-service with value-based payments, showing that effective incentive alignment hinges on using value metrics focused on the tool's quality—measured through documentable developer effort—rather than downstream outcomes. Interestingly, we find that as development costs rise, the minimum fee-for-service required to align incentives actually falls.

Our framework offers practical guidance for policymakers and payers aiming to craft reimbursement models that support the development and diffusion of high-quality technologies. By aligning incentives for developers and providers, these models can help integrate cutting-edge tools into clinical practice, enhancing both quality and access.

4 - Paving the Way for More Accessible Cancer Care in Low-Income Countries with Optimization

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

Cancers are a growing cause of morbidity and mortality in low-income countries. Geographic access plays a key role in both timely diagnosis and successful treatment. In areas lacking well-developed road networks, seasonal weather events can lengthen already long travel times to access care. Expanding facilities to offer cancer care is expensive and requires staffing by skilled medical professionals, which are often in short supply. In this article, we propose a mathematical model to improve geographic access to cancer care by jointly considering expansions to care facilities and improvements to the road network. We model this as a multi-period stochastic facility location network design problem. In each period, a decision maker must simultaneously choose a set of facilities at which to add tertiary cancer services and a set of roads to improve while facing demand and travel time uncertainty. Once demand for cancer care and weather events are realized, patients observe road conditions and use the transportation network to travel towards the closed facility with available cancer services. We create a new path-based formulation of this problem and develop a new branch-price-and-cut algorithm with acceleration techniques that take advantage of this formulation's structure. We demonstrate our approach using Rwanda as a case study and show that the reductions in travel time to cancer care that are directly attributable to the road network improvements can be as high as 1 hour.

MA43

Building B Level 3 B303

Optimization and ML solutions for healthcare decisions

Invited Session

Health Applications Society

Chair: Kimia Ghobadi, Johns Hopkins University, Baltimore, MD, United States

1 - How Close are We to Global Optimality in Cancer Radiotherapy Optimization? Benchmark Datasets and Algorithms for Measuring the Gap

Gourav Jhanwar, Memorial Sloan Kettering Cancer Center, New York, NY, United States, Mojtaba Tefagh, Qijie Huang, Linda Hong, Hai pham, Vicki Taasti, Seppo Tuomaala, Anthony Magliari, Michael Folkerts, Ying Zhou, Jie Yang, Saad Nadeem, Masoud Zarepisheh

Cancer radiotherapy treatment planning involves solving complex, large-scale optimization problems to determine patient-specific machine settings—such as the angle, intensity, and shape of radiation beams—that maximize tumor targeting while minimizing damage to healthy tissues. These problems are often non-convex and combinatorial, and the research community has developed a range of effective heuristic methods to solve them within strict clinical time constraints. More recently, AI-based approaches have shown promise in further improving planning efficiency and quality. Despite these advancements, two critical questions remain largely unanswered: 1) how close are the solutions obtained by existing methods—heuristic or AI-based—to the global optimum? and 2) how do different algorithmic approaches compare against each other in a standardized and reproducible setting?

To address this gap, we introduce **PortPy** (Planning and Optimization for Radiation Therapy in Python), an open-source framework that provides curated datasets derived from real patient cases, along with a suite of baseline algorithms ranging from classical heuristics to exact optimization techniques and modern AI-based methods. By offering a unified platform for evaluating and benchmarking algorithms, PortPy promotes transparency, reproducibility, and interdisciplinary collaboration in radiotherapy optimization research.

2 - Hybrid Risk Score Optimization Models for Patient Falls

Emmett Springer, Johns Hopkins University, Baltimore, MD, United States, Fardin Ganjkanloo, Erik Hoyer, Daniel Young, Kimia Ghobadi

Risk scores are prevalent in healthcare settings, and are simple classification models that allow clinicians to perform a fast and easy risk prediction with a linear combination. However, most risk score assessments are done manually, which is cumbersome and prone to human error. We design a set of risk score models with logistic loss to predict the risk of patients and stratify them into k classes. We augment and compare the models with varying levels of knowledge integration into the models for their accuracy, robustness, and computational expenses. We benchmark our models to predict the risk of patient falls in hospital settings, using data from three Johns Hopkins Health System hospitals. Our results show that all models improve compared to the current standard-of-care with up to a 29% increase in the area under the precision-recall curve (from 0.48 to 0.62) for the hybrid knowledge-based models. This translates to a 14.1% increase in specificity (0.71 to 0.81) with the same sensitivity and up to a 5.6% increase in sensitivity (0.88 to 0.93) with the same specificity. The hybrid knowledge-base models perform best in robustness, interpretability, and practical adaptation, while black-box models outperform other models in accuracy for the available dataset. Our models' insights highlight the critical importance of impaired mobility and cognitive impairments in predicting fall risk. This approach provides a robust foundation for health systems to systematically enhance inpatient fall prevention protocols and patient safety using data-driven optimization techniques, contributing to improved risk assessment and resource allocation in healthcare settings.

3 - An Augmented Mip Model for Home Care Scheduling

Jing Liu, Johns Hopkins University, Baltimore, MD, United States, Kimia Ghobadi

We develop a rolling-horizon Predict-Act-Learn framework for mixed-integer decision making under uncertainty, particularly for problems that arise in the healthcare operations domains. At each horizon, Predict estimates workload and disruption risk; Act solves a robust mixed-integer program to produce a base schedule; during execution, Learn uses a reinforcement-learning to refine the plan in response to realized conditions, and update models and uncertainty sets for subsequent horizons. This iterative process is designed to improve the reliability of the solutions when realizations deviate from expectations. We test this process for a home-care scheduling

problem with our partner collaborator (Vivia Cares; a home-care provider in Hawaii). We track and measure key performance metrics in provider-client matching, task assignment, patient preference, provider availability, and consistency in provider-client relationships, among others. We then develop a nominal optimization model to improve efficiency while balancing various tradeoff metrics in the problem, and adapt the Predict-Act-Learn approach for this problem to enable better adjustments to unforeseen changes and realization of the assumption.

4 - Using Large Language Models in Medical Analysis Tasks

Kimia Ghobadi, Johns Hopkins University, Baltimore, MD, United States, Felix Parker, Nimeesha Chan, Chi Zhang

Medical analysis tasks are often complex, multidimensional, and multimodal. In this talk, we explore the use of large language models (LLMs) in performing medical tasks. We particularly consider multimodal LLMs, and propose a hybrid foundation model (TsLLM) that integrates time series high-frequency signals with natural language models. This framework employs specialized techniques, including a patch reprogramming layer that aligns time series embeddings with the LLM's representation space, multiple covariate handling strategies for complex physiological relationships, and contextual prompting with patient-specific information. We evaluate TsLLM on various medical tasks, including electrocardiogram signals and outcome metrics.

MA44

Building B Level 3 B304

New Frontiers in Finance, Operations, and Risk Management

Invited Session

MSOM: iForm

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

Co-Chair: Zhanzhi Zheng, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

1 - Solving Mixed Strategy Equilibrium Using Machine Learning: Evidence from Stock Exchange Fees

David Wu, Cornell University, Ithaca, NY, United States, Mao Ye

We study two-dimensional competition where both price and quality can vary continuously. In these models, only mixed-strategy Nash equilibria exist in the 2D space, as there is no single "best" product - only products that better align with individual consumer preferences. We exemplify this with stock exchanges, which compete for order flow by setting "make" and "take" fees. We develop an algorithm to solve for a mixed strategy equilibrium using neural networks that easily generalizes to any mixed strategy setting. Using the new method, we establish that competition increases overall market functioning through increased transactions. However, the majority of the increased surplus goes to one side of the market.

2 - Privacy-Enhanced Payment Systems

Brian Zhu, Columbia University, New York, NY, United States, Agostino Capponi, Michael Lee

Technological innovations enable digital privacy, but pose fundamental conflicts between freedom and control. We study the design of privacy-enhanced payment systems, valued for legitimate transactions but vulnerable to illicit financial activities. We distinguish two dimensions of privacy: identity privacy (masking the owner of an account) and transaction privacy (confidentiality in activity). Identity-private systems are suboptimal, enabling illicit operations to scale by distributing activity across multiple accounts. The optimal provision of transaction privacy is complex and tailored to the technological capabilities of illicit actors, but adopts a simple design when considering robustness to innovations in these technologies. Our framework generates a ranking of payment systems; public blockchains, despite abuse of identity-privacy, can dominate legacy systems under strong enforcement capabilities.

3 - Supply-Chain Finance: A Sharing Contract via the Return Function

David Yao, Columbia University, New York, NY, United States, Liao Wang, Shaohui Zheng

We develop a new "return function" R , as an alternative to the classical net profit function H of the newsvendor (NV) model. As random variables, R and H are equal in expectation (mean); in this sense, R can also be viewed as net profit. Yet, we show that R dominates H in concave order, and has other more desirable properties, e.g., increasing in the (realized) revenue. Based on the return function R , we propose a "sharing contract," between a producer and an investor, as an alternative to the traditional loan contract. That is, instead of the producer getting a loan from the investor, the producer and investor now share the realized value of R . The sharing contract is analyzed in detail for various objectives, including the traditional NV objective (linear), the risk measure CVaR, and a general concave utility. In all cases, it is shown that R always dominates H , in terms of increased benefits for both the producer and the investor. We also derive a real-time hedging strategy for the investor to mitigate the risk and compensate for profit loss in low-demand scenarios.

4 - Targeted Automation and Sustaining Human-AI Learning

William Schmidt, Emory University, Atlanta, GA, United States, Christina Imdahl, Kai Hoberg

In many decision processes, a planner must review and optionally adjust the recommendations that are generated by a decision support system (DSS). When the DSS is well-tuned to its task, adjustments by a planner can be rare and may even degrade the DSS's performance. Targeted automation could address these inefficiencies by predicting whether a planner will adjust a recommendation and improve performance of the system. The remaining recommendations can be automated. However, as more recommendations are automated, fewer will receive planner input. This may starve the prediction model of the observations it needs for retraining. Using four years of procurement ordering data from our research partner, a large materials handling equipment manufacturer, we develop and train machine learning algorithms that predict individual instances in which a planner will improve a DSS-generated procurement decision on order quantities. We mitigate the performance erosion that automation engenders by structuring the selection of the ML model's classification threshold similar to a newsvendor problem that balances the costs and benefits of under or over automating and accounts for the value of learning. Our approach automates around 84% of all DSS recommendations while retaining three times more improvements than random automation. Our research contributes to a broader debate on the allocation of decision authority between humans and algorithms, and creates a framework for targeted automation in financial and operational settings that balances the benefits of automation versus algorithmic learning.

MA45

Building B Level 3 B305

Empirical Operations in the Healthcare Value Chain

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Xinyu Liang, INSEAD, Singapore, Singapore

Co-Chair: Jun Li, Ross School of Business, University of Michigan, Ann Arbor, MI, United States

1 - Corporate Venture Capital and Operational Performance of Medical Device Startups

George Ball, Operations and Decision Technologies, Kelley School of Business, Indiana University, Bloomington, IN, United States, Moyan Li, Haizhen Lin

Medical device startups are a pivotal dimension of new product innovation in the U.S. healthcare industry. Numerous innovative medical devices originate from startups that rely on two primary sources of external funding: independent venture capital (IVC) and corporate venture capital (CVC). CVC firms bring unique industry knowledge and incentives to startups that may impact startup's operational performance

outcomes. Using difference-in-difference models and an instrumental variable strategy, we study the effect of CVCs on the operational outcomes of medical device startups. We find that CVCs foster more startup product approvals, but once launched, these products experience worse product quality, indicating a double-edged sword to CVC funding. Post-hoc analyses reveal that this negative quality effect for startups is accentuated when CVC firms lack prior experience in the startups' product market, but is not impacted by the CVC firm's own product quality problems.

2 - Preannounced Regulatory Inspections: FDA Oversight and Drug Quality Risk

Zach Wright, Marriott School of Business, Brigham Young University, Provo, UT, United States, John Gray, George Ball, In Joon Noh

When inspecting a U.S. drug manufacturing facility, FDA inspectors arrive unannounced. However, in foreign countries, FDA inspectors provide significant advanced notice of their inspection, often many weeks or even months before arriving. We examine the impact of these differing FDA inspection approaches on inspection outcomes and drug quality. We first use data from unannounced U.S. and preannounced Indian FDA inspections. Using a matching analysis to mimic the FDA's model for selecting facilities for inspection, we find no significant difference in the odds of unannounced U.S. and preannounced Indian facilities receiving the most severe inspection outcome, implying relatively equivalent drug quality. Next, we examine previously unexplored data from two unannounced inspection pilots conducted by the FDA in India. These pilots enable us to disentangle whether the equivalent inspection outcome result is indicative of similar drug quality or a greater difficulty of detecting drug quality issues with preannounced inspections. We find that the odds of a facility receiving the most severe inspection outcome are significantly higher for unannounced inspections compared to preannounced inspections at the same facility in India. Finally, using a matched sample, we demonstrate potential consequences of preannouncing inspections in India, showing that the number of serious adverse events for drugs made in Indian facilities that experience preannounced inspections are significantly greater than for drugs made in matched facilities in the U.S. that experience unannounced inspections. Our findings provide the first quantitative empirical evidence that preannouncing FDA inspections may allow drug manufacturers to operate with greater and undetected quality risk.

3 - From Waste to Worth: Bridging the Gap Between Surplus and Shortage Outside the Traditional Pharmaceutical Procurement Channel

Ji Won Moon, Pennsylvania State University, University Park, PA, United States, Hui Zhao, Suvrat Dhanorkar

Drug *surplus* and *shortage* represent byproducts of inefficiencies in pharmaceutical supply chains that could help alleviate one another. Using proprietary data from an online B2B platform serving as an alternative, direct procurement channel for surplus drugs—most of which are *short-dated* (i.e., expiring within a year)—we empirically examine the benefits and caveats of this emerging marketplace for buyers (e.g., pharmacies, hospitals, providers) and sellers (i.e., manufacturers), particularly focusing on the dynamics between short-dated products and drug shortages as they influence sourcing patterns for medically necessary drugs in this unique environment.

4 - Impacts of a Non-Profit Intermediary on Reliability of Drug Supply

Junghye Lee, University of Notre Dame, Notre Dame, IN, United States, Jennifer Ryan, Hyoduk Shin, Daewon Sun

Generic drug shortages pose a growing and costly challenge in the U.S. This paper examines the impact of non-profit intermediaries, a recent entrant alongside traditional for-profit entities like group purchasing organizations, on the reliability of a shortage-prone generic drug supply chain. We develop an economic model relevant for mature off-patent generics with a limited supplier base. The entry of a non-profit introduces a new, potentially more reliable supplier. Our analysis reveals that while non-profit entry enhances supply reliability for its members, it does not guarantee overall supply chain improvement. We also find that increasing drug reimbursement rates may not necessarily improve supply reliability. Furthermore, our research provides guidance for non-profit organizations, highlighting that their effectiveness in mitigating shortages depends heavily on implementation strategies, including the degree of vertical integration. These findings offer critical insights for policymakers, as reimbursement increases are often proposed to address shortages.

5 - Changing Standards and Drug Shortages in the Pharmaceutical Industry

Ivan Lugovoi, Kuehne Logistics University, Hamburg, Germany, Enno Siemsen

Drug shortages can pose significant health risks; the failure of the market to fulfill the demand for life-saving drug treatments can endanger the health of patients. Using a proprietary dataset, we examine the relationship between changes in the quality standards of drugs and drug shortages through the lens of competition. We encode US-Pharmacopeia monograph changes to document changing quality standards in drug markets. Our analysis establishes that relaxing standards leads to an increase in the level of competition; however, tightening those standards does not lead to a similar decrease in competition. Further, we find a non-linearity in the relationship between the number of competitors and the risk of drug shortages. When that number is low, any new entrant will significantly decrease margins while not adding sufficient slack in the market capacity to absorb the failure of any incumbent to supply, leading to an increased drug shortage risk. However, as the number of competitors increases, the impact on profit margins is reduced; market shares of individual manufacturers decrease; and the reserve capacity in the market can more readily absorb the impact of a shortage by any individual competitor, decreasing the shortage risk again.

MA46

Building B Level 3 B306

Advances in Experimentation and Personalization in Managerial Settings

Invited Session

Decision Analysis Society

Chair: Spyros Zoumpoulis, INSEAD, Fontainebleau, France

Co-Chair: Zeyi Chen, INSEAD, Singapore, Singapore

1 - Minimax Optimal Exploration for Nonparametric Multi-Arm Bandits

Brian Cho, Cornell Tech, New York, NY, United States, Nathan Kallus

We study the task of fixed confidence pure exploration in multi-arm bandits, which aims to minimize the sample complexity for finding a δ -correct answer for the exploration task (e.g. find the best arm). While pure exploration problems have primarily been studied from a theoretical perspective, the empirical performance of such methods suffer from overly conservative stopping times, restrictive parametric assumptions, and poor empirical performance. In this work, we provide a general approach for pure exploration tasks in the nonparametric setting where arms only satisfy boundedness and mean stationarity. From a theoretical perspective, we show that our setting is no harder than stationary Bernoulli bandits, and provide a novel anytime valid sequential test that (i) achieves asymptotic minimax optimality and (ii) ensures δ -correctness for a wide range of data generating processes. To improve the practical performance of our approach, we leverage asymptotic approximations of our testing method to reduce conservativeness and provide a batch schedule that maintains theoretical guarantees. Both on synthetic and real-world data, our approach drastically outperforms existing methods in the literature.

2 - Just Ramp-up: Unleash the Potential of Regression-Based Estimator for A/B Tests Under Network Interference

Qianyi Chen, Tsinghua University, Beijing, China, People's Republic of, Bo Li

A/B tests on modern platforms, including social networks and two-sided marketplaces, often encounter challenges from network interference. Extant research on causal inference under network interference based on single experiments often incurs substantial bias, particularly with complex interference. In this paper, we demonstrate the statistical benefit of merging the data of multiple experiments with different treatment proportions. Sequential experimentation with increasing traffic is often known as the ramp-up process in tech companies and is commonly applied due to operational reasons such as risk management and cost control. Beyond operational considerations, we find that regression-based estimators trained with such merged data achieve substantial bias reduction even with simple randomization schemes. Since the ramp-up process

is standard practice, our methodology does not require additional experiment resources, providing a true plug-and-play solution that is easy to implement. We present a rigorous theoretical analysis of the bias and variance for the linear regression estimator of the global average treatment effect (GATE) under general linear network interference. We show that bias plays a dominant role in practical settings, and we highlight the bias reduction achieved by merging data from experiments with different treatment proportions. We provide an intuitive explanation for the bias reduction and, through the same lens, illustrate the synergistic effect of pooling experiments and cluster-level randomization. Furthermore, we consider a more advanced estimator based on graph neural networks (GNN). Through extensive simulation studies across novel and challenging scenarios, we show that the regression-based estimator benefits remarkably from our methodology, achieving outstanding statistical performance.

3 - Dynamic Local Average Treatment Effects

Ravi Sojitra, Stanford University, Stanford, CA, United States, Vasilis Syrgkanis

We enable identification, estimation, and inference for Local Average Treatment Effect (LATE) estimands in multiple time period settings with noncompliance and allow for treatment dynamics. Dynamics occur when treatment is encouraged in each time period depending on previous encouragements, treatments, and states (e.g. short term outcomes and time varying covariates). Although one may hope to leverage estimates of the effects of switching from one sequence of treatments to another, such quantities are not estimable under noncompliance and standard identifying assumptions for dynamic settings and LATEs. We introduce two conditions to enable identification of Dynamic LATEs that quantify effects for subpopulations who would comply with encouragements. First, we show that One Sided Noncompliance enables identification of all Dynamic LATEs corresponding to treating in a single period only. Second, further adding Staggered Adoption enables identification of effects of treating in multiple time periods. In general, the second result holds when the treatment effect of not continuing to comply is uncorrelated with whether one continues complying. We also show that a sequential extension of Monotonicity in Imbens and Angrist (1994) is not sufficient for the first result and an additional assumption is necessary for the second result. Finally, we use the automatic debiased machine learning framework to perform plug-in estimation and inference based on our identification results.

MA47

Building B Level 3 B307

Assorted Topics III

Contributed Session

Contributed

Chair: Amir Abdollahi, Northeastern University, 38 River Park Street, Boston, 02494, United States

1 - Closed-Form Formulas for Cluster Sizing for Two-Level Hierarchical Networks with Source Routing

Eric Rosenberg, Seton Hall University, South Orange, NJ, United States

In a two-level hierarchical network, nodes are grouped into clusters. A node in one cluster sees the entire topology of that cluster, but only a summarized view of other clusters. With source routing, the source node of a connection does an initial route computation based on its view of the network, and additional route computations are performed as the connection makes its way towards the destination. We study the problem of choosing the number of clusters to minimize the total complexity of all the path computations required to compute a path from a source node to a destination node. We show that, under mild conditions, for all sufficiently large networks two-level hierarchical routing is superior to flat network routing. Utilizing the duality theory of geometric programming, we provide a closed-form solution estimate for the number of clusters to minimize the total routing complexity. We also provide a closed-form expression for a lower bound on the minimal total complexity. We illustrate the modelling assumptions for a variety of networks. Computational results for a range of network sizes and parameters show that our solution estimate yields near-optimal results.

2 - Customer Relationship Management for Intangible Goods: the Effects of Customer Responsiveness, Engagement, and Reactance on Customer Rejection

Ameya Badve, City University of New York, New York, NY, United States, Ramakrishnaiah Bezawada, Karthik Sridhar

ABSTRACT

In a market driven by technology and information, the consumption of intangible goods has gone up. Customer relationship management for intangible goods is quite different from tangible goods. This article introduces a conceptual framework of customer relationship management for intangible goods that captures customer relationship using a novel construct called customer rejection and hypothesizes the effects of customer commitment, customer engagement, customer responsiveness, and customer reactance on customer rejection. The authors empirically test the hypotheses using data from a mobile app consumption. They find that while the risk of customer rejection is mitigated by the higher levels of customer commitment, customer engagement, and customer responsive, such risk is increased with a higher level of customer reactance. However, the increased risk of customer reactance can be managed using active customer engagement and customer responsiveness. Finally, exploring the different types of customer rejection, the authors provide strategies for hard and soft customer rejections based on product usage and customer lifetime. This research provides theoretical and managerial insights into the customer relationship management for intangible goods.

3 - Predicting Customer Lifetime Value and Segmenting Clients Using Machine Learning in the Payment Industry

Reza Maihami, Mercer University, Atlanta, GA, United States, Ehsan Ahmadi, Priya Gandhi, Maria Gomez

In the rapidly evolving payment industry, understanding customer value is critical for optimizing resource allocation and long-term growth. This article develops a machine learning-based framework to predict Customer Lifetime Value (CLTV) using historical merchant data in a non-contractual, continuous interaction setting. Leveraging the BG/NBD and Gamma-Gamma models, we estimate future purchasing frequency and monetary value for each customer. The predicted CLTV values are then used to segment customers via clustering analysis, uncovering distinct behavioral profiles. The resulting clusters—Diamond, Gold, Silver, and Bronze—highlight significant disparities in revenue contribution and engagement patterns. These insights enable targeted marketing strategies and personalized engagement efforts, empowering payment companies to retain high-value customers and uplift lower-tier segments. Our approach demonstrates the power of data-driven decision-making in customer relationship management and strategic planning.

4 - Modeling Willingness to Collaborate for Resource Sharing in Food Bank Networks

Rashik Siddiquee, North Carolina State University, Raleigh, NC, United States, Julie Ivy, Osman Ozaltin

The food bank network consists of numerous independently operated partner agencies that support food-insecure populations using donations from both individual sources and the central food bank. While collaboration among these agencies, through the sharing of resources and information, has the potential to enhance overall network resilience, such collaboration is often informal, unstructured, and highly variable across agencies. The degree to which an agency engages in this collaborative behavior is influenced by internal capacity, trust, technological capability, and organizational culture. This study seeks to quantify each agency's willingness to collaborate (WTC). We introduce a data-driven framework that models WTC as a behavioral parameter, estimated using regression and machine learning methods on observable features such as supply stability, communication practices, and historical cooperation patterns. By explicitly modeling willingness to collaborate, this research contributes a novel behavioral dimension to humanitarian logistics and offers insights for designing stronger, more cooperative food distribution networks. It also performs a case study on the agencies affiliated with the Food Bank of Central and Eastern North Carolina that serve Wake County.

5 - Decoding the Code: Revealing Block Usage Patterns for Design- and Context-Sensitive Computational Thinking (CT) Assessment in Game-Based Constructionist Learning

Amir Abdollahi, Northeastern University, Boston, MA, United States

Computational thinking (CT) is foundational in contemporary education, and game-based constructionist learning (GBCL) has proven effective in fostering CT, especially through block-based programming (BBP). However, the creative nature of game design complicates evaluating student-created artifacts. While automated CT metrics are helpful, they often overlook nuances in BBP. We analyzed 483 Scratch projects to uncover coding patterns, model artifacts' genre, and assess their combined impact on CT

outcomes. We found biases in the way students explore and use Scratch blocks and identified two clusters of coding patterns, strongly correlating with game genres (design aspect), demographic factors (e.g., Scratch experience, gender), and CT scores (learning aspect). Visual block categories (e.g., Looks) were key in decoding game genres, while logical categories (e.g., Data) were essential in clustering coding patterns and predicting CT uptake. These findings present opportunities for design- and context-sensitive CT assessments, paving the way for adaptive CT assessments in GBCL curricula.

MA48

Building B Level 3 B308

Human-AI Collaboration for Decision-Making

Invited Session

Decision Analysis Society

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

Co-Chair: Shixiang Woody Zhu, Carnegie Mellon University, Pittsburgh, PA, United States

Co-Chair: Holly Wiberg, Carnegie Mellon University, Pittsburgh, PA, 15213, United States

1 - Human-Guided Generate-Then-Optimize: a Unified Framework for Decision Refinement and Uncertainty Quantification

Minxing Zheng, Carnegie Mellon University, Pittsburgh, PA, United States

We introduce a Human-Guided Generate-Then-Optimize framework that integrates human preference feedback into decision-making pipelines under parameter uncertainty and distributional shift. Building on the predict-then-optimize paradigm, we model uncertainty over objective parameters with a generative model and obtain human preferences over candidate decisions. We first test for distributional shifts between human feedback and historical data; when shifts are detected, we refine the generative model via reward-based updates. Our framework improves both decision alignment and the efficiency of the uncertainty set of estimated parameters under distributional shift. This provides a unified approach for decision refinement, uncertainty-aware optimization, and adaptive alignment to human preferences in dynamic environments.

2 - The Value of AI Advice: Personalized and Value-Maximizing Ai Advisors are Necessary to Reliably Benefit Experts and Organizations

Nicholas Wolczynski, UT Austin, Austin, TX, United States, Maytal Saar-Tsechansky, Tong Wang

Despite advances in AI's performance and interpretability, AI advisors can undermine experts' decisions and increase the time and effort experts must invest to make decisions. Consequently, AI systems deployed in high-stakes settings often fail to consistently add value across experts and organizations and can even diminish the value that experts alone provide. Beyond harm in specific domains, such outcomes impede progress in research and practice, underscoring the need to understand when and why different AI advisors add or diminish value. To bridge this gap, we stress the importance of assessing the value AI advice brings to real-world contexts when designing and evaluating AI advisors. Building on this perspective, we characterize key pillars -- pathways through which AI advice impacts value -- and develop a framework that incorporates these pillars to create reliable, personalized, and value-adding advisors. Our results highlight the need for value-driven development of AI advisors that advise selectively, are tailored to experts' unique behaviors, and are optimized for context-specific trade-offs between decision improvements and advising costs. They also reveal how the lack of inclusion of these pillars in the design of AI advising systems may be contributing to the failures observed in practical applications.

3 - Llm Feedback on Science: a Large-Scale Randomized Field Experiment

Binglu Wang, Northwestern University, Evanston, IL, United States, Weixin Liang, Jiahui Xue, Yuhui Zhang, Hancheng Cao, Dashun Wang, Yian Yin

The transformative impact of Large Language Models (LLMs) across various sectors has demonstrated their potential to enhance productivity and facilitate complex decision-making processes, significantly altering innovation workflows. Despite their widespread adoption in commercial and consumer domains, the integration of LLMs into the unique processes of scientific discovery and innovation has been minimally explored. In this study, we present evidence from a large-scale randomized field experiment involving 34K recent preprints authored by 50K scholars across eight STEM fields and 136 global regions. We find that LLM-generated feedback significantly influences the revision of preprints and encourages the continued use of LLMs in academic writing. Notably, the impact of LLMs varies by region and domain expertise, highlighting both the promise and the inequality in their adoption. These findings suggest that, when thoughtfully deployed, LLMs can help democratize access to high-quality feedback and support more inclusive innovation practices on a global scale.

4 - Using Machine Learning to Diagnose Dynamic Human Decision Making: Insights from a Bioinnovation Context

Ziqian Ding, Carnegie Mellon University, Pittsburgh, PA, United States, Zhaohui Jiang, Param Singh

In various policy and business settings, humans are making high-stakes, dynamic decisions by managing trade-offs over time. Meanwhile, there are advancements in machine learning algorithms that are designed to optimize actions explicitly considering dynamic trade-offs. This paper introduces a framework demonstrating how these ML algorithms can be leveraged to diagnose and improve human decision-making in complex, dynamic settings. We implement this framework in a biomedical innovation context, where bioscientists make resource allocation decisions to maximize scientific success across a large number of projects over a span of 16 years. These decisions involve balancing immediate gains (exploitation) with the potential for learning across projects (exploration) in each period. We train a Combinatorial Neural Upper Confidence Bound (CN-UCB) model and use it as a benchmark to diagnose the decision-making processes of scientists. To understand the sources of inefficiencies, we estimate a dynamic structural model of human behavior, capturing deviations from the algorithm along two key dimensions: exploration-exploitation strategies and project success predictions. Interestingly, we find that scientists exhibit higher prediction errors than our algorithm, and tend to over-explore---they attempt too many projects they are unfamiliar with. Besides, humans on average allocate fewer trials to a project, and thus they are less likely to "delve deep" into a project than the algorithm. Finally, we find that unilaterally improving one of the two types of inefficiency can not fully recover the improvements of the algorithm over humans, emphasizing the importance of correcting both prediction errors and suboptimal explorations in diagnosing dynamic decision making.

5 - Validating LLM-as-a-Judge Systems in the Absence of Gold Labels

Luke Guerdan, Carnegie Mellon University, Pittsburgh, PA, United States

The LLM-as-a-judge paradigm, in which a judge LLM system replaces human raters in making decisions about the quality of other generative AI (GenAI) system outputs, plays a critical role in scaling and standardizing GenAI evaluations. To validate such judge systems, evaluators assess human--judge agreement by first collecting multiple human ratings for each item in a validation corpus, then aggregating the ratings into a single, per-item gold label rating. In many cases, however, items or rating criteria may be vague or ambiguous, or admit multiple interpretations, so a given rater (human or LLM) may deem multiple answers "reasonable" or "correct." We call such rating tasks indeterminate. Problematically, many indeterminate rating tasks rely on forced-choice elicitation, whereby raters are able to select only one answer option for each item. In this paper, we introduce a framework for LLM-as-a-judge validation for indeterminate rating tasks. We draw theoretical connections between different measures of judge system performance under different human--judge agreement metrics, and different rating elicitation and aggregation schemes. We demonstrate that differences in how humans and LLMs resolve indeterminacy in forced-choice rating tasks heavily bias LLM-as-a-judge validations. Through extensive experiments involving 11 real-world rating tasks and 8 commercial LLMs, we show that standard validation approaches select judge systems that are highly suboptimal, performing as much as 30% worse than judge systems selected by alternatives we propose. We conclude with concrete recommendations for more principled and reliable approaches to LLM-as-a-judge validation.

MA49

Building B Level 3 B309

Decision Making in E-Commerce

Invited Session

MSOM: Supply Chain

Chair: Jussi Keppo, National University of Singapore, Singapore, Singapore

Co-Chair: Bretislav Hajek, National University of Singapore, Singapore, Singapore

1 - Same-Day OR Next-Day? Transparent TIME-Dependent Shipment Pricing for E-Fulfillment

Uta Mohring, Eindhoven University of Technology, Eindhoven, Netherlands, Melvin Drent, Ivo Adan, Willem van Jaarsveld

Fulfilling online orders quickly is increasingly vital in e-commerce, with many companies offering same-day shipment services to meet rising customer expectations. However, companies risk overpromising their shipment services when the offered shipment options and corresponding fees are not aligned with the ability to fulfill these orders, leading to delayed orders and customer dissatisfaction. Indeed, fulfillment centers may be unable to hand over all orders to the carriers by the agreed deadlines. We propose a transparent time-dependent shipment policy where shipment options and fees depend on the remaining time until the next shipment deadline but are consistent each day. Such a shipment policy can influence customers' shipment preferences to help balance workload while remaining marketable to customers. We build a parsimonious model of a stochastic e-commerce fulfillment center operating this transparent time-dependent shipment policy, where customers arrive according to a Poisson process. We present an exact steady-state performance analysis based on a periodic Markov chain model and establish structural properties of the optimal shipment policy. Motivated by its practical appeal, we propose the two-level time-dependent shipment policy that differentiates two fees for same-day shipment. Numerical results indicate that this shipment policy achieves near-optimal solutions and clearly outperforms benchmark policies prevalent in practice that rely on time-independent fees. This work provides actionable guidance for e-commerce companies to better balance demand for same-day and next-day shipment with the available capacity in their fulfillment centers. Our transparent time-dependent shipment policy, designed to be disclosed to customers, offers a simple and customer-friendly solution.

2 - Study on the Overlap of Finite Mixture Modeling of Categorical Sequences

Yingying Zhang, Western Michigan University, Kalamazoo, MI, United States, Xuwen Zhu

Categorical sequences play an essential role in many fields, but research, especially the cluster analysis of this type of data, is limited. In this paper, the overlap character of categorical sequences is quantified. The proposed methodology involves derivation and calculation of the exact overlap expressed as the sum of misclassification probability between two clusters. Based on the calculated overlap value, algorithms are designed to simulate clusters satisfying user-defined overlap. The application of the proposed method is achieved through internet users' web-browsing sequences. Customers' internet navigation patterns are recorded and classified into distinct clusters. By investigating the overlap feature among various clusters, the similarities and dissimilarities of web-browsing behaviors can be clearly observed, which gives valuable insights for platforms for targeted advertising or group-specific recommendations.

3 - Information Design for Supply Chains with Priority Suppliers

Rufat Ismayilov, ESMT GmbH, Berlin, Germany, Ozan Candogan, Huseyin Gurkan, Boxuan Zhou

This paper studies the problem of information design in a supply chain where a retailer faces two upstream suppliers and must privately communicate market demand information. One supplier holds a priority advantage in fulfilling orders over the other. The retailer's challenge is to design an information disclosure policy that accounts for this priority structure while minimizing the expected lost sales cost. We model the interaction between the retailer and the suppliers using a Bayesian persuasion framework, where the retailer privately sends demand signals to each supplier, who then decides on their production quantities. We explore how the retailer should optimally inform suppliers with different priority levels. Our analysis reveals that the retailer should strategically favor the supplier with lower priority by providing more informative signals, while keeping the higher-priority supplier relatively less informed.

4 - Ad Feeding on A Consumer-to-Consumer Media Platform

Xuying Zhao, Texas A&M University, College Station, TX, United States, Zheyin (Jane) Gu

In this study, we investigate a C2C media platform's strategy to maximize advertising revenue under the threat of self-posted misinformation. Our model captures content providers' strategic decisions to post content on the platform as well as consumers' decisions to view content on the platform.

5 - Inventory Management and Assortment Planning In An Apparel Closed-Loop Supply Chain

Heletjé Van Staden, UCD Michael Smurfit Graduate Business School, Monkstown, Ireland, Hannah Yee

We consider an apparel rental platform with a fixed membership pool and non-stationary demand with partial observability. The platform aims to optimize garment purchasing decisions to best meet customer demand while minimizing operational costs. We model the problem as an ambiguous partially observable Markov decision process, which allows us to account for both uncertainty regarding the underlying demand distributions governing the demand and ambiguity regarding the platform members' demand preferences. We use reinforcement learning to approximate the optimal inventory control and assortment decisions. For the instances considered, we find that the algorithm improves on the static order up to policy by 6%, on average. By adjusting the inventory position to changing demand, our policy is able to limit dissatisfaction among members at the expense of holding slightly more inventory. Our work contributes to the closed-loop supply chain literature and can be used to support rental platforms with inventory and assortment procurement decisions given ambiguous and uncertain trends in customer preferences.

6 - Strategic Enforcement of Noncompete Agreements with Rival Active Differentiation

Kunyang Li, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Wenhui Zhao

Talent loss due to poaching—where employees leave to join competing firms—poses multifaceted risks to firms, particularly the leakage of trade secrets and tacit knowledge. To mitigate these risks, many firms globally adopt noncompete agreements (NCAs). However, NCAs embody an inherent conflict between firms seeking to protect intellectual assets and employees' right to mobility, fueling ongoing debate over their adoption and regulation. Existing discussions on how should the mother-firm design and use NCA to restrict the flow of talents to its competitors to suppress competitions, while overlooking the significance and necessity of the joint development of talents career and industries. We develop a game-theoretical modeling framework to capture the interaction between a firm and a talent, and examine the enforcement strategy of NCAs with active rival differentiation. Our findings indicate that the departure of talents can incentivize poaching firms to pursue differentiation strategies. This reduces the scope of noncompete restrictions and thereby encourages firms to be more supportive of talent mobility. Our findings provide valuable insights for both firms and regulators, offering guidance on structuring NCAs to balance trade secrets protection and employee mobility, thereby fostering talent development and industry growth.

MA50

Building B Level 3 B310

Strategic Queueing

Invited Session

MSOM: Service Operations

Chair: Ricky Roet-Green, Simon Business School, University of Rochester, Rochester, NY, United States

1 - Equilibrium and Regulation in Strategic Queues via Stochastic Approximation

Ran Snitkovsky, Collier School of Management, Tel Aviv University, Tel Aviv, Israel

Regulating admission in non-Markovian queueing systems often requires simulation, as closed-form solutions are typically unavailable. This challenge is further compounded when customers act strategically—making join/balk decisions based on expected delays—thereby inducing a feedback loop between customer behavior and system performance. In this talk, I introduce stochastic approximation algorithms designed for such settings and demonstrate their effectiveness in two canonical regulation problems: (1) maximizing social welfare and (2) maximizing revenue. The proposed approach exploits the regenerative structure of the underlying queueing process, allowing for asymptotic performance guarantees under mild regularity conditions.

2 - Observable Multi-Class Priority Queues: Rational Abandonment Strategies

Philipp Afèche, University of Toronto, Rotman School of Management, Toronto, ON, Canada, Simon Laumer

We study two questions regarding rational customers' abandonment behavior in observable multi-class priority queues: (1) What are the characteristics of an equilibrium in customers' join/balk/abandonment decisions? (2) How do these characteristics depend on customer attributes?

We consider an observable multi-class M/M/1 priority queue with strict preemptive priorities. The service order among customers in the same class is first-come-first-served. The service rate is equal for all customers. Each customer belongs to one of p priority classes, where p is an arbitrary integer. Each class is characterized by its arrival rate, linear delay cost rate and service reward. The arrival processes are exogenous, but upon arrival, every customer is forward-looking and solves an optimal stopping problem to maximize their expected utility (expected reward minus delay cost). The customer decides whether to balk or join, and if they join, whether to stay or abandon at every state transition (service completion or customer arrival), considering the future expected system evolution.

We provide analytical sufficient conditions for the equilibrium abandonment behavior to have the following threshold structure: (i) Each class has a threshold on the queue position of the last customer in that class, such a customer abandons only if their position goes above this threshold. (ii) These thresholds decrease in the priority level of the class, i.e., higher priority has a (weakly) smaller threshold.

These sufficient conditions depend explicitly on key model parameters (offered loads, delay costs and rewards of the different classes) and provide intuition on the system economics.

3 - Dynamic Pricing and Release TIME Control for Service Systems

Kevin Shang, Duke University, Durham, NC, United States, Chen-An Lin, Peng Sun

This paper studies a mechanism design problem for a single-server queue where customers are served in a first-come-first-serve (FCFS) manner. Customers are heterogeneous and have private information about their valuation of immediate service and sensitivity to waiting time. Impatient (versus patient) customers are more sensitive to delays and place higher value on immediate service. The service provider offers a history-dependent menu, with each menu item specifying a price and release time for each type of customer, aiming to maximize the long-run average revenue. We show that the optimal menu depends on history only through the system's completion time and consists of four possible strategies: delaying patient customers, pooling all customers, screening customers (serving one type and rejecting the other), and rejecting all customers. In particular, it is optimal to delay the release time of patient customers beyond the current service completion time if and only if the completion time is below a certain threshold. This strategic delay enables the provider to charge a premium to impatient customers, thus increasing revenue. Notably, the delayed release time is set to a fixed value, independent of the current completion time, which simplifies implementation. Under the screening strategy, the decision to serve patient or impatient customers depends on a trade-off between the utility a customer derives from joining the queue and the negative externality caused by increased congestion. These findings highlight how timing and differentiated pricing can be used to optimize revenue in queues with strategic delays.

4 - When Strategic Customers Meet Strategic Servers: Individual and Social Optimization in Many-Server Queueing Systems

Ragavendran Gopalakrishnan, Smith School of Business at Queen's University, Kingston, ON, Canada, Yueyang Zhong, Amy Ward

We initiate the study of interactions between strategic customers and strategic servers in many-server queueing systems. In the seminal works of Naor (1969) and Knudsen (1972), arriving customers decide whether to join the system based on a trade-off between their reward from service and cost of waiting, while servers operate at fixed exogenous rates. The joining threshold (maximum queue length) induced by self-interested customers at equilibrium is higher than socially optimal (i.e., they "over-join"), a result known as Naor's inequality. This is because they fail to internalize the congestion externality they impose on other customers. Although Naor's inequality is robust across a wide spectrum of queueing models, its validity when servers are also strategic (in choosing their work-speeds) is unknown. Within a large-system asymptotic framework, we show that strategic servers provide slower than socially optimal service at equilibrium, thereby deterring customers from joining the system, and when this effect outweighs customers' tendency to over-join, Naor's inequality is reversed. Next, we propose a scheme that charges an entry fee from customers and offers a performance-based compensation for servers, which realigns individual incentives with the social optimum. Finally, we analyze how the overall welfare gains at the social optimum (relative to the equilibrium) are distributed across customers and servers, in order to unearth potential equity trade-offs in policy design.

MA51

Building B Level 3 B311

Multi-Agent Learning in Markets

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Negin Golrezaei, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Rigel Galgana, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Heterogeneous Multi-Agent Bandits with Parsimonious Hints

Amirmahdi Mirfakhhar, Umass Amherst, Amherst, MA, United States, Xuchuang Wang, Jinhang Zuo, Yair Zick, Mohammad Hajiesmaili

We study a hinted heterogeneous multi-agent multi-armed bandits problem (HMA2B), where agents can query low-cost observations (hints) in addition to pulling arms. In this framework, each of the M agents has a unique reward distribution over K arms, and in T rounds, they can observe the reward of the arm they pull only if no other agent pulls that arm. The goal is to maximize the total utility by querying the minimal necessary hints without pulling arms, achieving time-independent regret. We study HMA2B in both centralized and decentralized setups. Our main centralized algorithm, GP-HCLA, which is an extension of HCLA, uses a central decision-maker for arm-pulling and hint queries, achieving $O(M4K)$ regret with $O(MK\log T)$ adaptive hints. In decentralized setups, we propose two algorithms, HD-ETC and EBHD-ETC, that allow agents to choose actions independently through collision-based communication and query hints uniformly until stopping, yielding $O(M3K2)$ regret with $O(M3K\log T)$ hints, where the former requires knowledge of the minimum gap and the latter does not. Finally, we establish lower bounds to prove the optimality of our results and verify them through numerical simulations.

2 - Markets with Heterogeneous Agents: Dynamics and Survival of Bayesian vs. No-Regret Learners

Yoav Kolumbus, Cornell University, Ithaca, NY, United States, David Easley, Eva Tardos

We analyze the performance of heterogeneous learning agents in asset markets with stochastic payoffs. Our main focus is on comparing Bayesian learners and no-regret learners who compete in markets and identifying the conditions under which each approach is more effective. Surprisingly, we find that low regret is not sufficient for survival: an agent can have regret as low as $O(\log T)$ but still vanish when competing against a Bayesian with a finite prior and any positive prior probability on the correct model. On the other hand, we show that Bayesian learning is fragile, while no-regret learning requires less knowledge of the environment and is therefore more robust. Motivated by the strengths and weaknesses of both approaches, we propose a balanced approach to utilizing Bayesian updates that improves robustness and adaptability to distribution shifts, providing a step toward a "best-of-both-worlds" learning approach. The method is general, computationally efficient, and easy to implement. Finally, we formally establish the relationship between the notions of survival and market dominance studied in economics and the framework of regret minimization, thus bridging these theories. More broadly, our work contributes to the understanding of dynamics with heterogeneous types of learning agents and their impact on markets.

3 - Allocating Public Goods via Dynamic Max-Min Fairness: Long-Run Behavior and Competitive Equilibria

Chido Onyeze, Cornell University, Ithaca, NY, United States, Siddhartha Banerjee, Giannis Fikioris, Eva Tardos

Dynamic max-min fair allocation (DMMF) is a simple and popular mechanism for the repeated allocation of a shared resource among competing agents: in each round, each agent can choose to request or not for the resource, which is then allocated to the requesting agent with the least number of allocations received till then. Recent work has shown that under DMMF, a simple threshold based request policy enjoys surprisingly strong robustness properties, wherein each agent can realize a significant fraction of her optimal utility irrespective of how other agents' behave. While this goes some way in mitigating the possibility of a 'tragedy of the commons' outcome, the

robust policies require that an agent defend against arbitrary (possibly adversarial) behavior by other agents. This however may be far from optimal compared to real world settings, where other agents are selfish optimizers rather than adversaries.

Our work aims to bridge this gap by studying the existence and properties of equilibria under DMMF. To this end, we first show that despite the strong robustness guarantees of threshold based strategies, no Nash equilibrium exists when agents participate in DMMF using such strategies. On the positive side, however, we show that for the symmetric case, a simple data-driven request policy guarantees that no agent benefits from deviating to a different fixed threshold policy. Furthermore, we show that the resulting equilibrium outcomes from this strategy can be significantly better compared to what follows from the robustness guarantees.

4 - Learning to Bid in Forward Electricity Markets Using a No-Regret Algorithm

Dorsa Majdi, Stanford University, Palo Alto, CA, United States, Arega Getaneh Abate, Jalal Kazempour, Maryam Kamgarpour

It is a common practice in the current literature of electricity markets to use game-theoretic approaches for strategic price bidding. However, they generally rely on the assumption that the strategic bidders have prior knowledge of rival bids, either perfectly or with some uncertainty. This is not necessarily a realistic assumption. This paper takes a different approach by relaxing such an assumption and exploits a no-regret learning algorithm for repeated games. In particular, by using the \emph{a posteriori} information about rivals' bids, a learner can implement a no-regret algorithm to optimize her/his decision making. Given this information, we utilize a multiplicative weight-update algorithm, adapting bidding strategies over multiple rounds of an auction to minimize her/his regret. Our numerical results show that when the proposed learning approach is used the social cost and the market-clearing prices can be higher than those corresponding to the classical game-theoretic approaches. The takeaway for market regulators is that electricity markets might be exposed to greater market power of suppliers than what classical analysis shows.

MA52

Building B Level 3 B312

Current Topics in The Energy Transition

Invited Session

MSOM: Sustainable Operations

Chair: Christian Kaps, Harvard Business School, Cambridge, MA, United States

1 - Utilities's Managed Home Charging Programs for Electric Vehicles

Ali Fattahi, Johns Hopkins University, Baltimore, MD, United States

Experts estimate 20 million electric vehicles will be on U.S. roads by 2030, and the majority (around 80%) of the electric vehicle drivers will use home charging. Many utilities are designing managed home charging programs to centrally manage charging times to reduce cost, avoid new and aggravated peaks and blackouts, and ensure grid stability. These managed home charging programs are either active, in which the utility continuously controls the charging while the vehicle is plugged in, or passive, in which the participants decide when to charge based on preannounced low-rate episodes. We study jointly designing and executing these active and passive programs. We present a program-design model, which produces a menu of the charging programs, tailored for each driver type, and a load-management model, which dynamically manages the load supply to each individual participant. The load-management model consists of a large number of nonhomogeneous participants, and it is a large-scale mixed-integer nonlinear stochastic problem. We present an effective approximation method, conduct thorough theoretical and numerical analyses of our approximation, and provide worst-case bounds for its error components. Our methodology provides detailed insights on the amount and timing of the improvements achievable in cost and demand variability by offering managed home charging programs, and by customizing the passive programs. It also offers detailed insights on the significance of the tradeoff between cost and demand variability. We find promoting a culture of charging electric vehicles every night may significantly increase utilities' total cost if passive programs have high participation levels.

2 - Behavioral Generative Agents for Energy Operations

Omer Karaduman, Stanford University, Stanford, CA, United States, Kuang Xu, Cong Chen

Accurately modeling consumer behavior in energy operations remains challenging due to inherent uncertainties, behavioral complexities, and limited empirical data. This paper introduces a novel approach leveraging generative agents--artificial agents powered by large language models--to realistically simulate customer decision-making in dynamic energy operations. We demonstrate that these agents behave more optimally and rationally in simpler market scenarios, while their performance becomes more variable and suboptimal as task complexity rises. Furthermore, the agents exhibit heterogeneous customer preferences, consistently maintaining distinct, persona-driven reasoning patterns. Our findings highlight the potential value of integrating generative agents into energy management simulations to improve the design and effectiveness of energy policies and incentive programs.

3 - Selecting Battery Portfolios: the Value of Heterogeneity

Joonho Bae, University of Michigan Ross School of Business, Ann Arbor, MI, United States, Roman Kapuscinski, John Silberholz

Given rapid technological advancements and the growing availability of batteries, including retired batteries from electric vehicles, battery operators are faced with the opportunity to invest in heterogeneous systems, also known as "hybrid energy storage systems." These systems are constructed using battery modules of different types, ages, and capacity ratings. However, making this decision is complex due to various factors, such as battery degradation or energy losses, which significantly affect the projected benefits of a diverse battery portfolio. Specifically, this study focuses on contingency applications, such as backup battery power systems, where the battery operator commits to supplying energy during emergencies, in the presence of uncertainties. This paper uses analytical modeling to systemically characterize the benefits of heterogeneity in battery systems and establish general lessons.

4 - Two-Sided Subsidies for Electric Vehicles: the Role of Regional Characteristics

Saed Alizamir, University of Virginia, Charlottesville, VA, United States, Michael Blair, Runyu Tang

Electric vehicles (EVs) are touted as the future of urban transportation and a solution to climate change. Despite their benefits, EV adoption faces significant headwinds because of high prices, range anxiety, and limited charging infrastructure. To overcome these obstacles, many governments have introduced financial incentives that directly support EV purchasers or hybrid policies that subsidize both buyers and charging station developers. Given the global variation in policies, the key questions are what type of subsidy program is optimal, and how does the optimal program change with the local market characteristics. Answering these questions and understanding the benefits of localized hybrid subsidies is the focus of this paper.

We characterize the optimal hybrid subsidy mechanism and identify conditions under which incentivizing EV buyers alone is suboptimal. We find that the two subsidy instruments can be substitutes or complements, depending on the net benefit of EVs to society at large. Leveraging data from Norway, we calibrate key model parameters and estimate that policies optimized for each region can enhance social welfare by up to 25% compared with national-level policies. Our model offers policy recommendations on how legislators should adjust EV subsidies considering regional characteristics.

MA53

Building B Level 3 B313

Learning and Optimization in Revenue Management

Invited Session

Revenue Management

Chair: Ruihao Zhu, Cornell University, Ithaca, NY, United States

Co-Chair: Hanzhang Qin, National University of Singapore, Singapore, Singapore

1 - Contextual Online Decision Making with Infinite-Dimensional Functional Regression

Haichen Hu, Massachusetts Institute of Technology, Cambridge, MA, United States, Rui Ai, Stephen Bates, David Simchi-Levi

Contextual sequential decision-making is fundamental to machine learning, with applications in bandits, sequential hypothesis testing, and online risk control. These tasks often rely on statistical measures like expectation, variance, and quantiles. In this paper, we propose a universal algorithmic framework that learns the full underlying distribution, enabling a unified approach to all contextual online decision-making problems. The challenge lies in the uncountably infinite-dimensional regression, where existing contextual bandit algorithms all yield infinite regret. We innovatively propose an efficient infinite-dimensional functional regression oracle for contextual cumulative distribution functions (CDFs) and model every datum as a combination of context-dependent CDF basis functions. Our analysis reveals that the decay rate of the eigenvalue sequence of the design integral operator governs the regression error rate, and consequently, the utility regret rate. We can also recover the existing optimal rate in contextual bandits literature with finite-dimensional regression and so as exponential decay. We also provide a numerical method to compute the eigenvalue sequence of integral operators, enabling the practical implementation of our framework.

2 - How Humans Help LLMs: Assessing and Incentivizing Human Preference Annotators

Xiaocheng Li, Imperial College Business School, London, United Kingdom

Human-annotated preference data play an important role in aligning large language models (LLMs). In this paper, we investigate the questions of assessing the performance of human annotators and incentivizing them to provide high-quality annotations. We study it from a mechanism design perspective and propose methods that enjoy nice theoretical guarantees and can be easily implemented in practice.

3 - Market Design for Platform-Mediated Influencer Advertising

Yan Chen, Duke University, Durham, NC, United States, Sasa Pekec

We study a centralized market design for influencer-based digital advertising, where advertisers reach users only through intermediary content creators (influencers). The presence of intermediaries introduces two challenges in valuing advertiser-influencer matches: imperfect targeting, where advertisers must target all followers of a matched influencer, and multiple coverage, where a user may be reached through multiple matched influencers. We model this as a two-stage problem: first, the platform selects a subset of influencers to participate in the market; second, advertisers are matched with participating influencers. We design a greedy approximation algorithm with a worst-case performance guarantee of $(1 - e^{-2})/2$ for this NP-hard influencer selection problem. We also provide a linear representation of the market-clearing problem, which features a non-linear concave objective due to the subadditivity of multiple coverage in targeting, and establish a sufficient condition for the existence of an integral solution. Finally, we analyze platform-controlled influencer pricing, characterizing competitive equilibrium prices and their properties. We also propose a pricing scheme that is linear in the number of followers and prove its optimality under some conditions.

4 - Landscape of Policy Optimization for Finite Horizon Mdp's with General State and Action

Minda Zhao, Georgia Institute of Technology, Atlanta, GA, United States, Xin Chen, Yifan Hu

Policy gradient methods are widely used in reinforcement learning. Yet, the nonconvexity of policy optimization imposes significant challenges in understanding the global convergence of policy gradient methods. For a class of finite-horizon Markov Decision Processes (MDPs) with general state and action spaces, we develop a framework that provides a set of easily verifiable assumptions to ensure the Polyak-Kurdyka-Łojasiewicz (PKŁ) condition of the policy optimization. Leveraging the PKŁ condition, policy gradient methods converge to the globally optimal policy with a non-asymptotic rate despite nonconvexity. Our results find applications in various control and operations models, including entropy-regularized tabular MDPs, Linear Quadratic Regulator (LQR) problems, stochastic inventory models, and stochastic cash balance problems, for which we show an ϵ -optimal policy can be obtained using a sample size in $O(\epsilon^{-1})$ and polynomial in terms of the planning horizon by stochastic policy gradient methods. Our result establishes the first sample complexity for multi-period inventory systems with Markov-modulated demands and stochastic cash balance problems in the literature.

5 - Machine Learning-Assisted Upper Confidence Bound Algorithm for Multi-Armed Bandits

Wenlong Ji, Stanford University, Stanford, CA, United States, Yihan Pan, Ruihao Zhu, Lihua Lei

Multi-armed bandit (MAB) is a prevalent paradigm for sequential decision-making under uncertainty. However, traditional bandit algorithms often face a cold-start problem, as they rely on online data that becomes available only through interaction with the environment. At the same time, decision-makers can leverage machine learning models to make predictions on the true reward. While these predictions have the potential to increase the sample size, they could be biased due to distribution shift or imprecise learning, and hence cannot fully replace the true rewards. In this paper, we explore how to utilize predictions from black-box machine learning models to improve bandit learning. We introduce the Machine Learning-Assisted Upper Confidence Bound (MLA-UCB) algorithm, which provably reduces cumulative regret by effectively de-biasing the predictions. In particular, our method is adaptive to any unknown covariance matrix between true and predicted rewards. Through extensive simulations, we show that MLA-UCB can significantly outperform the classical UCB without predictions under various settings.

MA54

Building B Level 2 B216

Platforms, Algorithms and Retail

Invited Session

Revenue Management

Chair: Nil Karacaoglu, The Ohio State University, Columbus, OH, United States

1 - Expedited Service Promises: Evidence from a Food Ordering Platform with Self-Pickup

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

Food ordering platforms increasingly enable customers to order online for self-pickup. A key challenge in this business setting is how to provide customers with estimates of the order completion time. Such service promises can not only affect customers' ordering inclination (and thus customer acquisition) but also their satisfaction after order pickup (and thus custom retention), thereby impacting business performance. This research estimates the causal effects of changing such service promises on product sales and how the effects vary by order type and product characteristics.

2 - Impact of in-Session Price Increases in the Cart on Consumer Behavior: Evidence from An Online Marketplace

Ovunc Yilmaz, University of Colorado Boulder, Boulder, CO, United States, Domenico Fabrizi, Serkan Mehder, Jiankun Sun, Zifeng Zhao

Price adjustments in the cart during shopping sessions are often a part of the customer journey on online marketplace platforms. In this research project, we investigate how customers react to in-session price increases in the cart while they are still shopping, using click-level data. We leverage a regression discontinuity design (RDD) to causally estimate the impact of these price changes on purchasing decisions.

3 - Impact of Multi-Channel Adoption on Customer Purchasing Behavior During Supply Chain Disruption

Jiachen WEI, Michigan State University, East Lansing, MI, United States, Stanley Lim, Elliot Rabinovich, Timothy Richards

This study investigates consumer purchasing behavior and inventory management challenges during pandemic-induced disruptions, using detailed sales data from a multi-channel grocery retailer. We find that although the demand surge following stay-at-home orders (SHOs) is short-lived, its impact is substantial. Consumers exhibit intense bulk-buying behavior within the first two weeks, but quickly revert to pre-pandemic patterns thereafter. Furthermore, the adoption of online channels leads to significant channel-shifting effects, with a marked decline in offline purchases during the same period.

We highlight the risks of overlooking the short-term nature of the demand spike, as doing so may result in overstocking and missed sales opportunities. To address this, we propose a transshipment policy that rebalances inventory across stores within a region. Our results underscore the importance of responsive, flexible inventory strategies

that recognize temporal demand dynamics in the face of crisis-driven shocks.

4 - Modeling and Empirical Considerations in B2b Pricing

Wen Yun, Columbia University, New York, NY, United States, Assaf Zeevi, Gal Lex

This work reports on a collaboration with a global manufacturer, Corning, with the purpose of developing a pricing decision support system for salespeople in a B2B setting. Unlike B2C settings, B2B sales involve factors not captured in historical data—such as long-term client relationships and risk management—making it infeasible for algorithms to set prices directly, as is customary in B2C settings. Instead, algorithms provide price suggestions, while salespeople retain final authority for setting transaction prices. Our work addresses two key challenges that arise in this context. First, salespeople may only partially comply with algorithmic suggestions. We model a broad class of compliance behaviors and derive a near-optimal, interpretable price recommendation policy. Second, salespeople's use of private demand information introduces unobserved confounding in demand estimation. Using real transaction data, we apply causal inference methods with instrumental variables to estimate heterogeneous treatment effects. Combining the modeling and empirical results with key considerations in human–algorithm interaction, we design a practical decision support system. Counterfactual analysis of policy performance suggests revenue gains of 1.12%–1.85% when salespeople's compliance exceeds 60% and discount rates vary within 5%.

MA55

Building B Level 3 B314

Perspectives on Learning and Decision Making: Theory and Applications

Invited Session

Revenue Management

Chair: Wanning Chen, University of Washington, Seattle, WA, United States

1 - Emotionally-Aware Conversational Agents

Wanning Chen, University of Washington, Seattle, WA, United States, Yingfei Wang, Tongxin Zhou

We propose a novel approach for designing emotionally-aware conversational agents by integrating reinforcement learning (RL) and large language models (LLMs) within a Partially Observable Markov Decision Process (POMDP) framework. Unlike purely reactive conversational models, our chatbot dynamically tracks and anticipates latent user emotional states, strategically adapting its conversational strategies. By formulating the problem as a linear dynamical control task, we leverage Kalman filtering for state estimation and Thompson Sampling for adaptive reinforcement learning, enabling effective emotional state recovery and proactive dialogue management. Empirical evaluation demonstrates that our RL-enhanced chatbot significantly improves emotional alignment and conversational quality compared to baseline methods, closely mimicking behavior characteristics of high-quality human therapists. This forward-looking, adaptive methodology provides a theoretically grounded and practically effective framework for enhancing personalization, engagement, and emotional intelligence in AI-driven psychological counseling.

2 - Leakage-Robust Bayesian Persuasion

Mingda Qiao, MIT, Cambridge, MA, United States, Nika Haghtalab, Kunhe Yang

We introduce the concept of leakage-robust Bayesian persuasion. Situated between public persuasion and private persuasion, leakage-robust persuasion considers a setting where one or more signals privately sent by a sender to the receivers may be leaked. We study the design of leakage-robust persuasion schemes and quantify the price of robustness using two formalisms.

The first notion, k -worst-case persuasiveness, requires a scheme to remain persuasive as long as each receiver observes at most k leaked signals. We quantify the Price of Worst-case Robustness (PoWR $_k$)—the gap in sender's utility as compared to the optimal private scheme—as $\Theta(\min\{2^k, n\})$ for supermodular sender utilities and $\Theta(k)$ for submodular or XOS utilities, where n is the number of receivers. This result also establishes that in some instances, $\Theta(\log k)$ leakages are sufficient for the utility of the optimal leakage-robust persuasion to degenerate to that of public persuasion.

The second notion, expected downstream utility robustness, relaxes the persuasiveness and considers the impact on sender's utility when receivers best respond to their observations. By quantifying the Price of Downstream Robustness (PoDR) as the gap between the sender's expected utility over random leakage patterns as compared to private persuasion, we show that over several natural and structured distributions of leakage patterns, PoDR improves PoWR to $\Theta(k)$ or even $\Theta(1)$, where k is the maximum number of leaked signals observable to each receiver across leakage patterns in the distribution.

En route to these results, we introduce general-purpose algorithmic paradigms—subsampling and masking—for transforming private persuasion signaling schemes to leakage-robust ones.

3 - Uncertainty-Informed Decision Making for Intelligent Mobile Health Support

Asim Gazi, Harvard University, Cambridge, MA, United States, Bhanu Gullapalli, Daiqi Gao, Benjamin Marlin, Susan Murphy

Mobile health presents a new frontier in preventive medicine, where wireless devices are used to extend healthcare support beyond the clinic. Just-in-time adaptive interventions (JITAI) are precision mobile health systems that leverage wireless devices to estimate changes in a person's biobehavioral context and deliver timely, tailored interventions. JITAI decisions have traditionally relied on user queries (e.g., ecological momentary assessments) or reliable sensor measurements (e.g., global positioning system coordinates) to estimate changes in context, but recent advancements have leveraged machine learning-based predictions. A key challenge in doing so is that predictors are fallible, yet current JITAI learning and decision-making algorithms operate as if predictions are always perfect. Our work advances the state of the art by (1) developing methods to better quantify uncertainty in biosensor-informed predictions and (2) improving JITAI decision making with contextual bandits by leveraging measures of prediction uncertainty to improve decision making. In this talk, we will focus on our research in improving JITAI decision-making by leveraging measures of posterior prediction uncertainty. We draw inspiration from operations research, where uncertainty is fundamentally used to balance the risks of under- versus over-prediction. We demonstrate the value of our uncertainty-informed decision-making methods first in the setting of decision point scheduling for an oral health JITAI based on forecasts of user brush times. We then extend our work to the more complex setting of a contextual bandit selecting intervention actions based on predictions of biobehavioral context. We thereby demonstrate the value of uncertainty-informed decision making in improving the effectiveness of JITAI.

4 - Model-Free Meta-Bayesian Optimization via Offline Acquisition Function Learning

Alon Jacobson, Cornell University, Ithaca, NY, United States, Qian Xie, Peter Frazier

Bayesian optimization (BO) is a powerful methodology for globally optimizing expensive black-box functions, with applications ranging from hyperparameter tuning to materials design. Classical BO relies on probabilistic surrogate models, often Gaussian processes, paired with hand-crafted acquisition heuristics. However, performance can degrade when surrogate assumptions, such as smoothness or Gaussian distributions, misalign with the true objective. To address this, we propose a model-free, data-driven approach: learning acquisition functions directly from offline datasets of related optimization tasks. By formulating acquisition functions such as Expected Improvement (EI) and the Pandora's Box Gittins Index (PBGI) as loss minimizers, we train neural networks to approximate them end-to-end, without reliance on explicit surrogate models. In one-dimensional experiments, our learned acquisition functions match GP-based baselines in both optimization performance and loss, while higher-dimensional results reveal challenges of overfitting and training instability, pointing to directions for future work. This study highlights the promise of replacing hand-crafted models and heuristics in BO with learned, scalable, and domain-adaptable optimization policies.

5 - Provably Learning from Language Feedback

Wanqiao Xu, Stanford University, Stanford, CA, United States, Allen Nie, Ruijie Zheng, Aditya Modi, Adith Swaminathan, Ching-An Cheng

Interactively learning from observation and language feedback is an increasingly studied area driven by the emergence of large language model (LLM) agents. While impressive empirical demonstrations have been shown, so far a principled framing of these decision problems remains lacking.

In this work, we formalize the Learning from Language Feedback (LLF) problem, assert sufficient assumptions to enable learning despite latent rewards, and introduce transfer eluder dimension as a complexity measure to characterize the hardness of LLF problems. We show that transfer eluder dimension captures the intuition that information in the feedback changes the learning complexity of the LLF problem. We demonstrate cases where learning from rich language feedback can be exponentially faster than learning from reward. We develop a no-regret algorithm, called LLF-UCB, that provably solves LLF problems through sequential interactions, with performance guarantees that scale with the transfer eluder dimension of the problem. Our contributions mark a first step towards designing principled interactive learning algorithms from generic language feedback.

MA56

Building B Level 3 B315

Design and Analysis of Marketplaces

Invited Session

Auctions and Market Design

Chair: Chen Chen, New York University Shanghai, Shanghai, N/A

Co-Chair: Huseyin Gurkan, ESMT GmbH, Berlin, Germany

1 - Information Design of Sequential Resource Allocation

Shanlin Jiang, University of Texas at Dallas, Richardson, TX, United States, Shouqiang Wang

Efficient resource allocation under decentralized control and asymmetric information is central to many economic and intra-organization settings. This paper studies strategic dynamic information disclosure from a principal with proprietary access to demand information to an agent who controls the allocation of limited resources. We fully characterize the principal's optimal information policy and obtain novel prescriptive insights.

2 - Temporary Title

Pranav Nuti, University of Chicago, Chicago, IL, United States, Rad Niazadeh

Temporary Abstract

3 - Dynamic Learning for Joint Pricing, Advertising, and Inventory Management

Huseyin Gurkan, ESMT GmbH, Berlin, Germany, Bora Keskin, Rodney Parker

Startup firms often lack distinct functional departments like marketing and operations, making cross-functional decisions difficult. These interdependent choices are further complicated by the absence of historical customer data. Consequently, startups must simultaneously learn about customer preferences while making marketing (pricing and advertising) and operational (inventory) decisions. This paper studies a setting where a firm jointly determines pricing, inventory, and advertising over TT periods while learning about demand and advertising response models. We characterize sufficient conditions under which cross-functional decisions enable exponentially fast learning rates. Based on these conditions, we propose an easy-to-implement policy that is asymptotically optimal: its profit gap relative to a clairvoyant with perfect information grows only at order $\log T$. Numerical experiments show that jointly learning about advertising and demand models is critical for strong profit performance. Our findings highlight the importance of integrating marketing and operational decisions to improve business outcomes. The learning conditions we identify emphasize how the connections between these decisions are key to understanding consumer behavior. Moreover, numerical results show that avoiding experimentation to preserve short-term profits leads to significantly worse long-term performance. Overall, our integrated approach offers practical guidance for entrepreneurs and managers aiming for better resource allocation, improved customer insights, and enhanced profitability.

4 - Mechanism Design with Post-Allocation Inspection and Exclusion Penalties

Xuanjie Li, Duke University, Durham, NC, United States

We study repeated interactions between a seller and a buyer in settings where the seller can inspect the buyer's private type and potentially exclude the buyer from future rounds upon detecting misreporting. This additional "exclusion penalty" instrument can take different forms, such as deterministic or probabilistic inspection and exclusion, and leads to novel mechanism design challenges. We develop a unifying framework to study these forms of inspection and exclusion penalty, characterize optimal mechanisms, and provide a dynamic programming solution to implement the optimal mechanism in multiple rounds. Our results show how inspection capabilities, when paired with strategic exclusion decisions, can augment classical mechanism design, offer new levers to align incentives and improve revenues.

5 - Optimality of Public Persuasion for Single-Good Allocation

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

We study an information design problem in which a sender seeks to allocate an indivisible good to n receivers by strategically disclosing information. The good can be allocated to at most one receiver, and each strategic receiver decides whether to accept based on self-interest. Relevant applications include school advisors promoting students for job placements and incubators introducing startups to potential investors. The sender can either send different signals to different receivers (i.e., *private persuasion*) or broadcast the same signal to all receivers (i.e., *public persuasion*). After receiving signals, receivers may communicate with each other in their self-interest to further reduce uncertainty about the good. We demonstrate when the sender has a known preference among the receivers, public persuasion is optimal, regardless of how receivers communicate. The optimal public persuasion can be derived from a first-best relaxation problem that imposes only the receivers' participation constraints. We then focus on a special case where the good's characteristics are captured by a one-dimensional variable, and receivers' utility functions are linear in this variable. Using a dual approach, we derive optimality conditions for persuasion mechanisms. This leads to closed-form optimal mechanisms for the two-receiver case and an explicit characterization of the set of optimal mechanisms in the general case, thereby enhancing our understanding of their structural properties.

MA57

Building B Level 3 B316

HAS Distinguished Speaker Session 1

Invited Session

Health Applications Society

Chair: Pengyi Shi, Purdue University, West Lafayette, IN, United States

Co-Chair: Lauren Steimle, Georgia Institute of Technology, Atlanta, GA, United States

1 - Optimizing the Deployment of Defibrillators

Timothy Chan, University of Toronto, Toronto, ON, Canada

This talk discusses several projects on optimizing the deployment of defibrillators for cardiac arrest, including placement of static defibrillators and drone delivery. Practical impact will be presented along with a discussion of how these projects have opened new doors to other research topics.

2 - Shaping Health Policy with Operations Research: The Hepatitis C Experience

Jagpreet Chhatwal, Harvard Medical School, Mass General Hospital, Boston, MA, United States

Over the past 12 years, Jagpreet Chhatwal and his team have developed a suite of mathematical models and decision tools that have directly informed hepatitis C elimination strategies. These models have been used by the White House, the Centers for Disease Control and Prevention, and the World Health Organization to guide

national and global health policies. In this talk, Dr. Chhatwal will highlight how these models shaped real-world decisions and the lessons learned from collaborations with policymakers. He will also discuss strategies for building successful partnerships and ensuring that operations research reach the stakeholders who need them most. The session will provide practical insights for the INFORMS community on how to maximize the real-world impact and visibility of their research.

MA58

Building B Level 3 B317

Innovation and Automation in Manufacturing

Contributed Session

Contributed

Chair: Junayed Pasha, Gannon University, Erie, PA, United States

1 - Technological Modularity in Resource Redeployment

Yi Luo, IE Business School , Madrid , Spain

This paper aims to examine how technological modularity can affect resource redeployability. Building on real options theory, we argue that technological modularity can simultaneously enhance inducements to redeploy capacity-constrained resources when volatility reveals superior uses, and lower the internal adjustment costs in the redeployment process. This dual mechanism suggests that technological modularity will amplify the positive relationship between volatility and firm market value identified in prior work on resource redeployment. To test our hypotheses, we use cross-sectional Fama-MacBeth regression with a patent-based measure on a panel of U.S. public firms. By integrating the literature on modularity and resource redeployment, this study endogenizes inducements and costs of redeploying resources, demonstrating that the technology architecture plays an important role in redeploying resources internally, and that technological modularity is a critical boundary condition for the market valuation of flexibility.

2 - A Research Study of Optimized Vehicle Design Space Exploration Using Multiple Criteria Decision Making

KyongRyol Yoon, Hyundai Motors, Seoul, Korea, Republic of, Mauro Munerato

Recent vehicle development is highly dependent on data-driven decision making with categorized target and value in earlier step. This means that a vehicle should satisfy purpose-based performance goal, reliable battery safety and competitive low TCO. However in vehicle concept design development, this strategy could be difficult choice because of getting less data and unreasonable result prediction. To solve this problem, several Automotive OEM has researched and developed an exploration space analysis methodology. It was valuable approach but another issue was upcoming, a lot of concept design candidate were an obstacle of right decision making. In this study, Multiple Criteria Decision Making was introduced and researched in order to get an efficient exploration and optimum design. The result was shown that ranking and sorting MCDM algorithm with real-time weighting could be the key player in initial vehicle development.

3 - Capability OR Liability? the Role of Innovativeness Under Government Sanctions

Zhihong TIAN, The Hong Kong Polytechnic University, Hong Kong, China, People's Republic of, Di Fan, Chengyong Xiao

Previous literature widely regards firm innovativeness crucial for building a capability for coping with environmental uncertainties. However, it remains unclear whether innovativeness remains beneficial under increasingly prevalent geopolitical risks, such as foreign government sanctions targeting specific firms. Drawing on an integrated framework between Resource Dependence Theory and the Core Rigidity perspective, we conceptualize sanctions as disruptions that constrain targeted firms' access to critical external resources, including supply chains, technology, and markets. Given innovative firms' heightened reliance on specialized external inputs and established routines, we argue that innovativeness may paradoxically undermine the firm's responsiveness and increase vulnerability during sanctions. Using a sample of 101 Chinese listed firms included in the sanction list issued by the US International Trade Administration of the Department of Commerce, our difference-in-difference analysis finds that sanctions undermine the inventory and profitability performance of these firms. Furthermore, we find that firms with high-quality (indicated by patent citations) and broader (indicated by the scope of patents) innovation suffer more from the sanctions. These results demonstrate that the sanctions effectively cut off the supply chain support necessary for innovation. Our findings suggest that without supply chain backing, innovativeness's resilience against sanctions proves to be a fragile facade. We discuss the contributions to operations and supply chain management literature in the areas of innovation and public policy, and provide managerial and policy implications.

4 - Flexible vs. Dedicated Technology Choice in the Presence of Multiproduct Subscription Programs

Liling LU, Singapore Management University, Singapore, Singapore

In recent years, automotive manufacturers have started offering multi-car subscription programs that give customers access to a menu of car models in a certain fleet with possibility of switching from one model to another for an all-inclusive fee. The fleet's maintenance and depreciation costs during the program are borne by the manufacturers before these cars are sold in the preowned market. This paper studies how offering such a subscription program shapes the firm's flexible versus dedicated technology choice and capacity investment decisions. We consider a two-product firm that makes these decisions under uncertain demand in a single subscription period. The firm faces two types of subscription demand based on which product is requested at the beginning of the period. Halfway through the period, a proportion of the existing customers returns their allocated product and requests switching to the other product. The firm incurs a cost unless this request is satisfied by the existing fleet of products. In deciding the manufacturing volume the firm considers not only the realized demand but also the future switching requests where some products can be reserved to satisfy these requests. We characterize the firm's optimal decisions and examine, both analytically and numerically using experiments calibrated to data, how demand correlation shapes the firm's profitability with each technology as well as the optimal technology choice. We demonstrate that our findings can be significantly different from those in the traditional setting (where the firm sells its products) and we explain why.

5 - Minimizing Energy Consumption of a Robotic Arms by Considering Base and Idle Positions in Automated Warehouses

Seung Hyeon Eom, Kyung Hee University, Yongin-si, Korea, Republic of, Byung hun Song, Hyun Woo Jeon, Jin Goo Jeong

The rise of e-commerce has accelerated the deployment of warehouses supported by robotic arms, leading to increased energy consumption. In response, this study develops a model to estimate and minimize the energy consumption of robotic arms based on their operational movements within automated warehouse systems. Two key decision variables—the base position (BP) and the idle position (IP)—are determined to minimize energy usage. These variables are influenced by operational factors such as idle time, payload, and positional error. Using an L9 orthogonal experimental design, simulation experiments validate the significant impact of these factors on the robotic arms' energy consumption and optimal positioning. To guarantee the uniqueness of the global optimum, energy consumption is mathematically formulated as a strictly convex function with respect to BP. The simulation experiment results indicate that as the position error of objects on the conveyor belt increases, the impact of BP and IP control becomes more significant. The proposed energy estimation and optimization model contributes significantly to operational cost reduction and sustainability in automated warehouse environments.

6 - Automation in Part Number Stamping Operations: A Case Study

Junayed Pasha, Gannon University, Erie, PA, United States, Daniel Rutkowski, Jayed Pasha, Madalyn Myers, Andrea Martinez Ortuno

A stamping and inspection process is a critical step in various manufacturing lines, where parts are marked with identification numbers, checked for quality, and either passed to the next production stage or scrapped due to non-conformance. This research presents a comprehensive simulation modeling and analysis of this critical step, comparing manual and automated operations. A simulation model is developed under this study that replicates the flow of parts through the system, capturing key performance metrics, such as throughput, error rates, and cycle times. By creating digital twins of both the manual and automated versions of this processing step, the impact of automation on the overall system performance is quantified. A number of scenarios are explored in order to account for variability in part arrival rates, processing times, and defect occurrences. Simulation results indicate that automation significantly improves process reliability and throughput, with a significant increase in the number of successful parts processed per hour. Additionally, the automated system demonstrates a significant reduction in quality-related scrap rates. These

improvements have cascading effects on adjacent processing steps, increasing overall line efficiency by a significant amount. Hence, the insights gained from this research could guide manufacturers in making data-driven decisions about process automation and resource allocation.

MA59

Building B Level 3 B318

Healthcare Operations and Analytics

Contributed Session

Contributed

Chair: Mahyar Mahmoudi, Oklahoma State University, Stillwater, OK, United States

1 - Customizing Follow-up of Cancer Patients: Case of Breast Cancer in Morocco

Abdelaziz Berrado, EMI, Mohammed V University in Rabat, Rabat, Morocco, Loubna Benabbou

This ongoing research work, which involves artificial intelligence experts and medical oncologists is about the development of an intelligent system for a predictive prognosis and a personalized surveillance/follow-up of breast cancer patients. The system would save lives, lower the cost of treatment and follow-up and relieve the capacity of the oncology treatment centers for better management of new and old cases of breast cancer.

The targeted intelligent decision support system will guide the practitioner in the surveillance of patients with breast cancer based on the risk of relapse specific to each patient. It will enable the oncologist or general practitioner to introduce several attributes relative to the patient's history and condition, and in return obtain a personalized proposal for a follow-up plan for the patient, depending on the risk of relapse and the history of the patient. Follow-ups will vary in rhythm/frequency and also in tests to be prescribed.

The resulting system will, make it possible to optimize the surveillance of former patients and to propose a personalized monitoring plan instead of a systematic monitoring frequency as recommended by the current guidelines (every three months) . This will help avoid inappropriate monitoring that is either too heavy or too light and will allow better management of hospital capacity. Furthermore, a better prediction of the risk of relapse will make it possible to intervene more quickly for cases at high risk of relapse. The system will make it possible to delegate the surveillance task to non-specialist doctors.

2 - Gatekeepers' Competition and Optimal Payment Scheme in Healthcare

Ji-Su Lee, KAIST, Daejeon, Korea, Republic of, Woonam Hwang, Taesik Lee

This study investigates how different payment schemes influence the healthcare quality and referral decisions of lower-level hospitals operating as gatekeepers in hierarchical healthcare systems. We develop analytical models to examine how market structure—monopoly versus duopoly—and the intensity of competition affect the ability of payment mechanisms to align hospital behavior with socially optimal outcomes. We analyze several payment schemes, including fee-for-service, fund-holding gatekeeper, and outcome-based penalty models, and propose a novel referral threshold-based subsidy scheme. Our findings reveal that conventional fee-for-service schemes generally fail to achieve the social optimum due to their limited incentive alignment. In monopolistic markets, outcome-based penalties effectively induce optimal hospital behavior. In duopolistic settings, the effectiveness of payment schemes critically depends on competition intensity. Under moderate competition, multiple schemes can achieve optimal outcomes, though with varying cost implications for payers and profit outcomes for providers. However, when competition is intense, only the proposed referral threshold-based subsidies can restore socially optimal outcomes. This research offers actionable insights for policymakers seeking to design payment schemes that enhance the performance of gatekeeper-based healthcare systems by accounting for competitive dynamics and market structure.

3 - A Stochastic Optimization Framework for Prioritizing SDOH Documentation for Chronic Disease Patients

Ifeoluwa Solaru, Louisiana State University, Baton Rouge, LA, United States, Laura Ikuma

This study develops a stochastic optimization framework to tackle the operational challenges of social determinants of health (SDOH) documentation in resource-constrained healthcare systems using 49,492 chronic patient records from a Louisiana medical center between January and December 2024. The study formulates a chance-constrained programming model that maximizes expected clinical impact, quantified through effect sizes - Cramer's V and Phi ($\Phi=0.16$ for alcohol use, $\Phi=0.073$ for housing instability) - while probabilistically bounding documentation uncertainty across 12 social needs. The model incorporates a greedy ϵ -approximation algorithm that ranks needs by risk-adjusted yield ($\Phi / (\text{unknown rate} + \epsilon)$), identifying alcohol use (priority score = 0.625), housing instability (0.110), and social connection (0.093) as the highest screening priorities. Probabilistic constraints derived from refusal patterns (ranging from 89% decline for physical activity to 26% for alcohol use) and Monte Carlo validation demonstrate that the stability of the screening priorities is sustained under parameter uncertainty and varying response completeness. The framework requires only three inputs per social need -clinical effect size (Φ), documentation completeness ($1 - \text{unknown rate}$), and refusal probability, making it computationally tractable and highly adaptable. Sensitivity analysis indicates that the model remains robust to $\pm 15\%$ variation in effect sizes (95% CI) and maintains feasibility even when unknown response rates exceed 80%. This prioritization approach offers a reproducible, data-driven strategy to guide SDOH screening efforts, aligning limited clinical resources with population-level risk impact while accounting for real-world documentation limitations.

4 - Chronic Risk & Structural Shield: Diabetes, Income, and Cancer Outcomes in the United States

Deepthi Singh, California State University - Long Beach, Long Beach, CA, United States, Banafsheh Behzad

Cancer is a leading global cause of death, with incidence projected to rise by 70% in the coming decades. Modifiable risk factors—such as diabetes, obesity, tobacco, and alcohol use—contribute to nearly half of all cancer cases. Our study focuses on the role of diabetes in cancer outcomes and how income and health coverage influence this relationship. Using panel data from 50 U.S. states and the District of Columbia from 1997 to 2019, we estimate dynamic panel models via system Generalized Method of Moments (GMM). We combine internal lag instruments with external policy shocks to address endogeneity. States are grouped into three clusters by lifestyle patterns to examine heterogeneity. Our results indicate diabetes as a strong, consistent predictor of both cancer incidence and mortality. Its impact is significantly moderated by income, with higher income reducing diabetes-related cancer risk. We find that health coverage is positively associated with cancer incidence, likely reflecting increased detection, but it does not consistently reduce mortality. Cluster-specific results show that the income-buffering effect is strongest in states with low-risk lifestyles and higher coverage. The effect of diabetes on cancer is substantial and shaped by socioeconomic context. Income and access play critical roles in moderating this risk. These results underscore the importance of prioritizing income-sensitive, chronic disease-focused interventions, particularly in high-risk and underserved states, to reduce the cancer burden linked to diabetes.

5 - Variational Inference Federated Learning for Incomplete Multi-Modality Imaging on Alzheimer's Disease Prediction

Zhiyang Zheng, Georgia Institute of Technology, Atlanta, GA, United States, Ruiqi Chen, Yi Su, Kewei Chen, David Weidman, Teresa Wu, Shih-Chung Lo, Fleming Lure, Jing Li

Alzheimer's Disease (AD) is a devastating neurodegenerative disease, often preceded by Mild Cognitive Impairment (MCI), making the early prediction of MCI conversion to AD critical for timely intervention. While multi-modality neuroimaging enhances predictive accuracy, its widespread application in machine learning (ML) models is hampered by incomplete datasets across medical sites and significant privacy barriers to data sharing. Current federated learning (FL) solutions in neuroimaging typically focus on single-modality data, thereby missing the synergistic advantages of multi-modal approaches. This paper introduces FedVICML (Federated Variational Inference for Cross-Modality Learning), a novel FL framework specifically designed to tackle these challenges by enabling sites with varying access to imaging modalities to collaboratively train models without centralized data sharing, thus preserving patient privacy. FedVICML uniquely integrates FL with a novel cross-model variational inference methodology, employing two distinct Bayesian neural networks (one for multi-modal and another for single-modal data), and an extended variational inference-based mechanism to facilitate knowledge transfer between these models. The FedVICML framework provides a decentralized, privacy-preserving learning with incomplete multi-modality data by innovatively integrating FL and cross-model variational inference. We demonstrated its efficacy by applying the proposed framework to

simulated data and real-world data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) for predicting MCI to AD progression; FedVICML achieved notable gains in prediction performance compared to isolated FL baselines, underscoring its potential for robust, collaborative medical research under heterogeneous data conditions.

6 - Caim-Oird Risk Tool: Enhancing Clinician-Ai-Interaction to Monitor Opioid-Induced Respiratory Depression in Surgery and Trauma Patients Using a Multi-Level Interpretability Approach

Elham Nasarian, Virginia Tech, Blacksburg, VA, United States, Huaiyang Zhong, Kwok-Leung Tsui

Opioid therapy is commonly used for postoperative pain management but carries significant risks, particularly opioid-induced respiratory depression (OIRD), which is associated with increased mortality, longer hospital stays, and higher healthcare costs. Despite its prevalence, OIRD remains unrecognizable. This study compares the PRODIGY score with our own risk prediction model using interpretable ML techniques to identify key predictors of OIRD. The Random Forest model outperformed others in recall and precision, suggesting its potential for integration into clinical decision support systems (CDSS). Additionally, we developed a new risk score tool based on the most important factors identified. Our findings emphasize the need for better predictive tools, continuous monitoring, and timely interventions to reduce the clinical and economic burden of OIRD.

7 - Synthetic - Empirical Log Fusion for Deep Learning Prediction of Outpatient Queue Dynamics

Daesan Park, Pusan National University, Busan, Korea, Republic of, Dohee Kim, Hyerim Bae

Classical queueing formulas provide closed-form steady-state metrics and clear managerial insights into system sensitivity under stable loads; however, their analytical reach diminishes when demand spikes or service rules change. Discrete-event simulation (DES) accurately reproduces such dynamics, but repeated runs and calibration hinder real-time use. Recent deep learning predictors enhance speed, but most are trained only on synthetic DES episodes—introducing model bias—or solely on sparse field logs, which leaves them insufficiently exposed to rare transients and measurement noise.

To address these limitations, we propose a data fusion framework that provides an off-the-shelf deep learning predictor with both synthetic and empirical samples. A configurable simulator generates both routine and stress scenarios. Each episode delivers queue length paths and moments of waiting time, while the clinic logs contribute the same indicators under routine conditions. Within this framework, we systematically adjust the ratio of synthetic to empirical data to train the predictor and analyze how this ratio balances steady-state accuracy against responsiveness under stress. The proposed framework is validated using outpatient clinic data to demonstrate its practical applicability while still offering rapid inference for operational decision support.

Results demonstrate (i) how prediction error varies with log share across routine and stress conditions, (ii) that fusion training remains resilient to measurement noise and log sparsity, and (iii) offers practical guidance for selecting simulation-to-log ratios—uniting the strengths of analytic formulas, simulation, and deep learning.

8 - Nonlinear Risk Mapping and Conformal Confidence Intervals for Accurate Dr Prediction

Mahyar Mahmoudi, Oklahoma State University, Stillwater, OK, United States, Tieming Liu

Diabetic retinopathy (DR) remains a leading cause of vision loss worldwide, particularly among individuals with long-standing diabetes. This study introduces a novel framework to construct a clinically interpretable and computationally efficient risk index for early detection of DR using electronic health records. We selected ten essential predictors through a robust ensemble feature selection approach using gradient boosting, achieving a compact yet powerful predictive model. To discretize continuous features, we adopted a decision-tree-based method to determine optimal bin cutoffs, improving granularity and interpretability in clinical settings.

Unlike prior linear risk scoring methods, we contributed a nonlinear transformation to calculate risk scores, enabling more realistic risk weighting across variable ranges. Additionally, we applied conformal prediction to estimate confidence intervals for every possible total risk score, offering a robust alternative to traditional statistical approaches such as Wilson's method.

Our risk index model demonstrated strong performance with an AUC of 0.89 on the internal validation cohort and 0.83 on an external validation cohort, confirming its generalizability across independent populations. The model provides both probability scores and uncertainty bounds, supporting confident decision-making in DR screening initiatives.

This work advances the methodological foundation for risk score generation in chronic disease management, particularly in under-resourced environments where interpretability, ease of use, and confidence quantification are critical. Our approach bridges the gap between complex machine learning models and actionable clinical tools for proactive DR management.

MA60

Building B Level 3 B319

AI, Learning, and Games

Contributed Session

Contributed

Chair: Jinkun Lee, East Carolina University, Greenville, NC, United States

1 - On the Complexity of Causal Invariant Learning Under Distribution Shift

Haotian Xie, Ohio State University, Columbus, OH, United States, Simon Zhang, Kun Jin, Cathy Xia

Out-of-distribution (OoD) generalization remains a fundamental challenge in supervised machine learning. Models often fail when applied to data that differ from the training distribution. Traditional methods, which rely heavily on statistical correlations, are particularly vulnerable to such distributional shifts.

Causal AI offers a principled alternative by modeling the underlying cause-and-effect relationships in data using causal graphs. This talk focuses on causal invariant learning: a family of approaches that seek to uncover stable, causally relevant features by learning *invariant representations* that disentangle true causal structure from spurious correlations. These representations aim to remain predictive under both environmental interventions and distributional changes.

We will review key algorithmic strategies, including Invariant Risk Minimization (IRM), Variance Risk Extrapolation (V-REx), and other causally motivated objectives. While these methods show promising empirical results, a comprehensive theoretical understanding is still emerging.

In this talk, we explore the conditions under which causal invariant learning can succeed, and present preliminary results on its computational complexity and generalization guarantees. Our aim is to illuminate both the practical potential and theoretical boundaries of causal methods for robust machine learning in the face of distribution shift.

2 - Navigating Content Strategy in the Face of Platform Moderation: the Case of Youtube's Borderline Policy

Shervin Shahrokhi Tehrani, University of Texas at Dallas, Dallas, TX, United States, Reza Roshangarzadeh

Social media platforms like YouTube rely on algorithmic moderation to manage harmful content, often demoting so-called borderline content—material that skirts but does not explicitly violate platform guidelines. This study investigates how content creators dynamically adjust their strategies under such moderation uncertainty. Using a curated dataset of over 3,800 videos across sensitive topics—including election integrity, COVID-19, and the Ukraine war—we construct a borderline index by applying

Natural Language Inference (NLI) to quantify each video's alignment with known misinformation. We model creators' topic choices as a dynamic Bayesian learning process, where engagement outcomes act as noisy signals reflecting the platform's treatment of borderline content. Empirical results show that creators initially overestimate the engagement potential of borderline topics but recalibrate expectations downward with experience. We also document substantial spillover learning: creators revise beliefs about related, untested topics based on observed engagement shifts. Counterfactual simulations indicate that stricter moderation signals reduce borderline content prevalence and boost overall platform engagement. These findings shed light on how creators navigate opaque moderation environments and offer actionable insights for platform governance and content strategy. Our framework emphasizes the role of dynamic adaptation in content ecosystems and highlights how moderation policies can shape both content and engagement outcomes.

3 - Rs-Ort: a Reduced-Space Branch-and-Bound Algorithm for Optimal Regression Trees

Cristobal Heredia Galleguillos, University of South Florida, Tampa, FL, United States, Pedro Chumplitaz, Kaixun Hua

Mixed-integer programming (MIP) has emerged as a powerful framework for learning optimal decision trees. Yet, existing MIP approaches for regression tasks are either limited to purely binary features or become computationally intractable when continuous, large-scale data are involved. Naively binarizing continuous features sacrifices global optimality and often yields needlessly deep trees. We recast the optimal regression-tree training as a two-stage optimisation problem and propose Reduced-Space Optimal Regression Trees (RS-ORT)—a specialised branch-and-bound (BB) algorithm that branches exclusively on tree-structural variables. This design guarantees the algorithm's convergence and its independence from the number of training samples. Leveraging the model's structure, we introduce several tightening techniques—closed-form leaf prediction, empirical threshold discretisation, and exact depth-1 subtree parsing—that combine with decomposable upper and lower bounding strategies to accelerate the training. The BB node-wise decomposition further enables trivial parallel execution, further alleviating the computational intractability even for million-size datasets. On 11 UCI regression benchmarks containing both binary and continuous features, RS-ORT generally delivers lower training loss and superior test performance than state-of-the-art methods. Notably, on datasets with up to 2000000 samples with continuous features, RS-ORT can obtain competitive training performance with a simpler tree structure and provable optimality in four hours.

4 - Multimodal Personalized Federated Learning for Product Recommendation in Fast Fashion

Xinhui Yu, The University of British Columbia, Vancouver, BC, Canada, Sophie Liu, Chunhua Wu

Recommendation systems play a pivotal role in modern marketing and e-commerce, enabling businesses to personalize customer experiences, optimize product exposure, and enhance engagement and conversions. Building effective recommendation systems is particularly critical in dynamic, data-sensitive domains such as fast fashion.

We introduce pFedMMRec, a personalized federated learning framework for fast fashion product recommendation. Unlike conventional federated recommendation approaches that rely predominantly on item ID embeddings and often overlook rich multimodal data that influences purchasing decisions, pFedMMRec employs a novel dual-branch architecture. This structure separately models each customer's design preference and price sensitivity by leveraging multimodal product data, including images, text descriptions, and price information. To enhance the model's discriminative capabilities, we introduce a threshold-based regularization loss that creates greater separation between each customer's preferred and non-preferred design elements. Additionally, we develop personalized adapters for image and text processing that enable the model to learn customer-specific item representations while effectively leveraging collective knowledge across the federated network.

We validate our methodology through comprehensive simulation experiments conducted on four diverse clothing datasets spanning different product categories (long sleeves, pants, dresses, and outerwear) extracted from real-world fast fashion data. Experimental results demonstrate that pFedMMRec consistently outperforms existing methods, confirming the effectiveness of our multimodal, privacy-preserving, personalized approach in addressing the challenges of both limited information exploration and data heterogeneity in federated fashion recommendation systems.

This research contributes to the intersection of recommendation systems, marketing analytics, and federated learning by offering a scalable solution for adaptive, secure, personalized product recommendations in fast-paced retail environments.

5 - Managing the Learning-Retention Trade-off: a Field Experiment on Incentivizing Educational Content in a Video Game

Alexander Jensen, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Sebastien Martin, Karen Smilowitz

The integration of educational content into digital environments presents a fundamental trade-off between promoting learning and sustaining user engagement. Although educational technology has been studied extensively, educational video games, particularly those blending structured learning with open-ended play, remain underexplored in the operations management literature. Using gameplay data from a video game designed for children aged 5–12, we analyze how players alternate between “work” (educational content) and “play” (exploration, quests, and combat). We observe that user retention declines sharply after initial exposure to educational content and that early educational engagement is strongly correlated with permanent user exit. While these patterns are not necessarily causal, they highlight a critical behavioral tension in game design that is also applicable in broader educational settings. To address this, we are conducting a randomized controlled trial (June 2025) to test whether in-game notifications, delivered at different points during a session, can increase educational engagement without reducing retention.

6 - Optimization of Insulin Pump Infusion Rate by Reinforcement Learning

Jinkun Lee, East Carolina University, Greenville, NC, United States

Minimizing insulin dosage while effectively controlling plasma glucose concentrations is crucial for reducing the risk of insulin resistance and ensuring effective long-term diabetes management. We focus on designing an insulin pump infusion controller modeled by a system of ordinary differential equations that captures the dynamic interaction between glucose and insulin in the human body. The primary objective is to maintain plasma glucose levels within a healthy range over time. To achieve this, we employ Reinforcement Learning (RL) and investigate the structure of a weighted reward function to evaluate RL's ability to discover an optimal infusion strategy that meets predefined clinical targets. The RL-generated insulin infusion profiles will be compared to those derived from a Proportional-Integral-Derivative (PID) controller and current best practices. This comparison will be conducted by analyzing the area under the glucose concentration curve for each approach, to assess the suitability of RL for this application.

MA61

Building B Level 4 B401

Maintenance Optimization for Energy and Mobile Systems

Invited Session

Quality, Statistics and Reliability

Chair: Jeffrey Kharoufeh, Clemson University, Clemson, SC, 29621, United States

Co-Chair: Ahmed Aziz Ezzat, Rutgers University, Piscataway, NJ, United States

1 - Mobile Condition-Based Maintenance with Heterogeneous Repair Types and Durations

Emma Gibbs, University of Pittsburgh, Pittsburgh, PA, United States, Lisa Maillart, Oleg Prokopyev

This work considers the problem of performing condition-based maintenance on a set of geographically distributed assets via a single mobile maintenance resource. By formulating a Markov decision process (MDP), optimal timing and selection of positioning and repair decisions are determined dynamically as a function of asset health and maintenance resource location to minimize total downtime, travel, and maintenance expenses. As an extension of the previous work, assets may have heterogeneous non-unit repair durations, and maintenance resource has access to multiple repair actions. These generalizations facilitate the examination of how asset repair durations impact preventive maintenance of nearby assets, repositioning, idling in anticipation of asset deterioration, and tradeoffs between short-but-costly or long-but-inexpensive repairs. Optimal policies for the MDP are explored analytically and numerically with respect to graph structure and repair durations.

2 - Condition-Based Production for Stochastically Deteriorating Systems

Collin Drent, Eindhoven University of Technology, Eindhoven, Netherlands, Melvin Drent, Joachim Arts

Production systems degrade stochastically with use and may ultimately fail, incurring significant costs during scheduled maintenance. This degradation is generally influenced by the system's production rate: Higher production yields more revenue but also accelerates wear. As a result, production should be dynamically adjusted to balance revenue generation and system degradation between maintenance moments. In this context, we study a two-level decision-making problem. At the operational level, we investigate how to optimally control production to manage degradation between planned maintenance moments. At the tactical level, we study when to schedule these maintenance moments, assuming optimal operational control in between. We cast the operational problem as a continuous-time Markov decision process, and we prove that the optimal policy exhibits intuitive monotonic properties. We also provide sufficient conditions under which so-called bang-bang production policies, where production switches between extremes, are optimal. For the tactical problem, we identify a structural property that simplifies its optimization considerably. Numerical experiments show that our approach outperforms existing state-of-the-art methods in realistic scenarios.

3 - Optimal Replacement of Fixed-Base and Floating Offshore Wind Turbines

Morteza Soltani, Clemson University, Clemson, SC, United States, Jeffrey Kharoufeh, Amin Khademi

Floating offshore wind (FOW) turbines have emerged as a promising solution for harnessing wind energy in deep ocean waters where fixed-base turbines are infeasible. However, replacement of major components is a challenging challenge due to harsh environmental conditions and limited availability of specialized resources, such as jack-up vessels (JUVs). Unlike fixed-base turbines, FOW turbines can be towed to either shallow waters or shore for replacement activities. We present a framework for optimizing replacement decisions for both fixed-base and FOW turbines under stochastic environmental conditions. The environment states in shallow and deep waters are modeled as a bivariate Markov chain that exhibits (weak) positive correlation at each time step. A Markov decision process (MDP) model is created to jointly optimize replacement and JUV requesting decisions over an infinite planning horizon. The model incorporates stochastic dependence arising from degradation processes influenced by correlated environment states, and economic dependence due to sharing of a JUV across turbine types. We characterize the structural properties of the optimal policies and show that both replacement and JUV requesting decisions exhibit threshold behavior with respect to turbine degradation. Finally, we provide sufficient conditions under which these policies are monotone in the environmental state.

4 - Towards Foundational Models for Diagnostics and Optimization in Energy Systems

Murat Yildirim, Wayne State University, Detroit, MI, United States, MOHAMMAD BADFAR, Iman Kazemian, Paritosh Ramanan, Ratna Chinnam

Modern data-driven systems face two central challenges: the complexity of data engineering and its connection to decision-making. In this work, we present two complementary frameworks that address these bottlenecks—ADEPT and ATOM—each targeting automation and scalability in predictive analytics and constrained optimization. ADEPT (Automated Data Engineering Pipeline via Text Embeddings) eliminates the need for traditional feature engineering by converting raw multivariate time series into text-based representations and processing them through a variational bottleneck over pretrained embeddings. This approach offers resilience to missing values, irregular timestamps, and corrupted records, and enables end-to-end prediction pipelines with minimal manual intervention. Experiments in a hydropower dataset demonstrate ADEPT's consistent outperformance over conventional methods. In parallel, ATOM (Attention-based Time-series Optimization Model) offers a paradigm shift for condition-based maintenance by directly learning decision policies from raw sensor streams—circumventing the typical two-stage predict-then-optimize paradigm. ATOM integrates operational constraints, such as resource limits and crew schedules, through attention-masking, producing interpretable, cost-efficient maintenance policies. On commercial Li-ion battery fleets, ATOM achieves up to 35% cost savings over standard baselines, while demonstrating early-stage foundation model capabilities through transferability across cost regimes. Together, ADEPT and ATOM form an integrated vision for the future of automated, constraint-aware decision systems: one that fuses representation learning with decision-making and removes the traditional divide between data preparation and optimization.

5 - A Stochastic Optimization Framework to Integrate Turbine Failure Risks and Accessibility Forecasts to Offshore Wind Farm Operations, Maintenance, and Logistics

Muhammet Ceyhan Sahin, Wayne State University, Detroit, MI, United States, Feng Ye, Ahmed Aziz Ezzat, Onur Kilic, Murat Yildirim

This paper introduces ALBATROS (accessibility-aware logistics based approach for turbine reliability and operations under stochasticity), a novel optimization framework for coordinating operations, maintenance, and vessel logistics in offshore wind (OSW) farms. ALBATROS captures key sources of uncertainty through (i) sensor-driven prognostic models for turbine degradation and failure risks, (ii) data-driven met-ocean forecasts of wind and wave conditions for accessibility, (iii) stochastic modeling of electricity prices and wind-dependent power production, and (iv) detailed vessel routing and logistics for geographically dispersed farms. These components are unified in a two-stage stochastic mixed-integer model that jointly optimizes condition-based maintenance, vessel routing, and production planning. The model operates in a receding horizon fashion to coordinate hourly and daily decisions over a 60-day planning window. A comprehensive computational study involving 100 turbines across four OSW locations benchmarks ALBATROS against conventional practices. Results show significant improvements in cost reduction, turbine uptime, and system reliability. The study also quantifies the value of sensor-driven condition monitoring over periodic maintenance, demonstrates the role of accessibility forecasting in improving scheduling feasibility, and evaluates the impact of vessel availability windows. By integrating real-world data, forecasting, and logistics in a unified decision-making framework, ALBATROS provides a robust, data-driven solution for OSW O&M planning under uncertainty.

MA62

Building B Level 4 B402

Tensor-based Modeling and Control for Digital Manufacturing

Invited Session

Quality, Statistics and Reliability

Chair: Jianjun Shi, ISyE Georgia Tech, Atlanta, GA, United States

Co-Chair: Zihan Zhang, Georgia Institute of Technology, Atlanta, GA, United States

1 - Nonlinear Spatio-Temporal Run-to-Run Control: Leveraging Diffusion Models for Disturbance Management

Jianjun Shi, ISyE Georgia Tech, Atlanta, GA, United States

Run-to-run (R2R) control is widely used in advanced manufacturing processes. Conventional R2R control methods often assume that system disturbances exhibit linear structures or low-rank approximations in high-dimensional data, such as images or videos. While these assumptions facilitate implementation, they limit the effectiveness of R2R control in managing sophisticated manufacturing processes with intricate nonlinear disturbances. To address these challenges, we propose a novel nonlinear spatio-temporal R2R control framework that estimates system disturbances using a diffusion model and applies control actions to compensate for them. The proposed approach integrates an offline modeling phase to capture disturbance dynamics and an online control phase that dynamically adjusts actions to minimize deviations from the target system response, ensuring system stability and precision. Unlike conventional deterministic control strategies, the proposed method explicitly quantifies uncertainty, enabling more robust and informed decision-making. The effectiveness of this method is validated through simulation studies and a case study, demonstrating its adaptability in complex, high-dimensional environments.

2 - On Model Compression for Neural Networks: Framework, Algorithm, and Convergence Guarantee

Bo Shen, New Jersey Institute of Technology, Newark, NJ, United States, Chenyang Li, Jihoon Chung, Mengnan Du, Haimin Wang, Xianlian Zhou

Model compression is a crucial part of deploying neural networks (NNs), especially when the memory and storage of computing devices are limited in many applications. This paper focuses on two model compression techniques: low-rank approximation and weight pruning in neural networks, which are very popular nowadays. However, training NN with low-rank approximation and weight pruning always suffers significant accuracy loss and convergence issues. In this paper, a holistic framework is proposed for model compression from a novel perspective of nonconvex optimization by designing an appropriate objective function. Then, we introduce NN-BCD, a block coordinate descent (BCD) algorithm to solve the nonconvex optimization. One advantage of our algorithm is that an efficient iteration scheme can be derived with closed-form, which is gradient-free. Therefore, our algorithm will not suffer from vanishing/exploding gradient problems. Furthermore, with the Kurdyka-Łojasiewicz

(KL) property of our objective function, we show that our algorithm globally converges to a critical point at the rate of $O(1/k)$, where k denotes the number of iterations. Lastly, extensive experiments with tensor train decomposition and weight pruning demonstrate the efficiency and superior performance of the proposed framework.

3 - Regularized Tensor Completion for Structural Health Monitoring Data Imputation

Shenghao Xia, University of Arizona, Tucson, AZ, United States

Missing sensor data is a ubiquitous issue in structural health monitoring (SHM), often resulting from sensor failures and restricting system effectiveness. SHM data, collected from spatially distributed sensors over time, typically exhibit intricate spatial-temporal (ST) patterns, which pose difficulties for conventional imputation methods, especially in scenarios involving prolonged or continuous data loss. To address these challenges, this study proposes a novel imputation method that explicitly captures ST dependencies and enables accurate recovery of critical structural characteristics in both the time and frequency domains. An efficient optimization algorithm is developed to support robust and scalable estimation. The proposed method is validated through a real-world case study, demonstrating superior imputation accuracy and robustness compared to existing benchmark approaches.

4 - Sparsest: Exploiting Data Sparsity in Spatiotemporal Modeling and Prediction

Junfeng Wu, Rensselaer Polytechnic Institute, Troy, NY, United States, Yinan Wang

Spatiotemporal data mining has a wide range of applications in modeling and predicting various complex physical systems (CPS), i.e., transportation, manufacturing, healthcare, etc. Among all the proposed methods, the Convolutional Long-Short Term Memory (ConvLSTM) has proved to be generalizable and extendable in different applications. However, ConvLSTM and its variants are computationally expensive, which makes it inapplicable in edge devices with limited computational resources. With the emerging need for edge computing in CPS, efficient AI is essential to reduce the computational cost while preserving the model performance. Common methods of efficient AI are developed to reduce model redundancy (i.e., model pruning, etc.). However, model redundancy is limited in spatiotemporal data mining as the embedded dependencies are complex and hard to capture. Instead, there is a fair level of data redundancy, which has been largely overlooked in existing research. Using the sequence of images as an example, (1) the informative pixels at each frame are usually spatially sparse (i.e., with a fixed background), and (2) the informative features in the first-order difference of two adjacent images are possibly sparse due to the slow temporal evolution. Therefore, we develop a novel efficient ConvLSTM that pioneeringly exploits the sparsity of informative features. More specifically, the sparse convolution and the Delta algorithm are incorporated to exploit the spatial sparsity at each single time step and the sparsity in the first-order difference over a sequence, respectively. The experiment demonstrates that our proposed methods can achieve comparable performance to the original ConvLSTM with reduced computational complexity.

5 - Disco: Disentangling Signal Restoration and Corruption Identification via Sample-Wise Latent Low-Rank Similarity

Yifeng Wang, Georgia Institute of Technology, Atlanta, GA, United States, Shancong Mou, Jianjun Shi, Chuck Zhang

With the rapid advancement of sensing technologies, diverse signals can now be collected during complex industrial processes, enabling more effective strategies for quality control through signal restoration and anomaly (corruption) identification. While numerous approaches have been developed for fine-grained anomaly detection, existing state-of-the-art methods often struggle to simultaneously achieve accurate anomaly identification and nonlinear signal restoration due to the inherent entangling effect. To address this critical challenge, we propose DISCO (*Disentangling Signal Restoration and Corruption Identification*), a novel method that generalizes Robust Principal Component Analysis (RPCA) to the nonlinear domain by utilizing an autoencoder network and introducing the sample-wise latent low-rank similarity. This regularization enables DISCO to decouple the underlying signal structure from additive corruptions effectively. Extensive evaluations, including simulation experiments, case study, and ablation analysis, demonstrate that DISCO significantly outperforms existing benchmarks in both fine-grained anomaly detection and nonlinear background restoration.

MA63

Building B Level 4 B403

Reliability, Robustness, and Uncertainty Quantification

Contributed Session

Contributed

Chair: Orkun Irsoy, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Improvement in Airline Operations Using Six Sigma Methodologies

Ramesh Bollapragada, San Francisco State University, San Francisco, CA, United States

The purpose of the paper is to investigate the financial and operational benefits when deploying six sigma methodologies in airline operations. Specifically, it intends to improve the service operations of the airline industry. The six sigma continuous improvement methodology i.e. DMAIC (define, measure, analysis, implementation, and control) is used in this study. Six sigma statistical analysis is used to measure the current performance of each critical metric involving the operational processes within international and U.S. domestic airline industry. Using six sigma level or the current sigma level (whichever is higher) as the benchmarking tool, this paper compares the compensation expenses caused by the variations in operational processes before and after the sigma level is achieved. We present a new approach to set an internal benchmark and determine the tolerance level, when measuring airlines (international, domestic) operational standards. The findings depict the difference between the current sigma level and the proposed six sigma level for each critical metric. Through the six sigma methodology, we provide an improvement in the following six critical passenger-defined quality metrics: accident rate, casualty rate, flight delay rate, flight cancellation rate, baggage irregularity rate, and denied boarding rate. An analysis of the annual financial savings on each of the above operational metrics is presented. The aggregate annual savings obtained across all the critical metrics is in the order of billions of dollars discussed.

2 - Anytime-Valid Inference in Adaptive Experiments: Covariate Adjustment and Balanced Power

Daniel Molitor, Cornell University, Ithaca, NY, United States, Samantha Gold

Adaptive experiments such as multi-armed bandits offer efficiency gains over traditional randomized experiments but pose two major challenges: invalid inference on the Average Treatment Effect (ATE) due to adaptive sampling and low statistical power for sub-optimal treatments. We address both issues by extending the Mixture Adaptive Design (Liang and Bojinov, 2024; MAD) framework. First, we propose MADCovar, a covariate-adjusted ATE estimator that preserves anytime-valid inference guarantees while substantially improving ATE precision. Second, we introduce MADMod, which dynamically reallocates samples to underpowered arms based on statistical significance, enabling more balanced statistical power across treatments without sacrificing valid inference.

Both methods retain MAD's core advantage of constructing confidence sequences (CSs) that remain valid under continuous monitoring and arbitrary stopping. Theoretical guarantees show our ATE estimators are unbiased and yield shrinking, valid CSs under mild regularity conditions.

Empirically, we validate both methods using simulations and real data. In simulation, MADCovar reduces CS width by up to 40% relative to MAD. In a large-scale political RCT with ~32,000 participants, MADCovar achieves up to 3–4x precision gains. MADMod improves statistical power and inferential precision across all treatment arms, particularly for suboptimal treatments, by reallocating samples based on a user-specified function. Simulations show that MADMod sharply reduces Type II error while preserving the efficiency benefits of adaptive allocation.

Together, MADCovar and MADMod make adaptive experiments more practical, reliable, and efficient for applied researchers across many domains.

3 - Automated Analysis of Experiments Using Hierarchical Garrote

Wei-Yang Yu, Georgia Institute of Technology, Atlanta, GA, United States, V. Roshan Joseph

In this work, we propose an automatic method for the analysis of experiments that incorporates hierarchical relationships between the experimental variables. We use a modified version of the nonnegative garrote method for variable selection which can incorporate hierarchical relationships. The nonnegative garrote method requires a good initial estimate of the regression parameters for it to work well. To obtain the initial estimate, we use generalized ridge regression with the ridge parameters estimated from a Gaussian process prior placed on the underlying input-output relationship. The proposed method, called HiGarrote, is fast, easy to use, and requires no manual tuning. Analysis of several real experiments are presented to demonstrate its benefits over the existing methods.

4 - Analysis and Optimization of Robustness in Multiplex Flow Networks Against Cascading Failures

Orkun Irsoy, Carnegie Mellon University, Pittsburgh, PA, United States, Osman Yagan

Networked systems are susceptible to cascading failures, where the failure of an initial set of nodes propagates through the network, often leading to system-wide failures. In this work, we propose a multiplex flow network model to study robustness against cascading failures triggered by random failures.

The model is inspired by systems where nodes carry or support multiple types of flows, and failures result in the redistribution of flows within the same layer rather than between layers.

To represent different types of interdependencies between the layers of the multiplex network, we define two cases of failure conditions: layer-independent overload and layer-influenced overload.

We provide recursive equations and their solutions to calculate the steady-state fraction of surviving nodes, validate them through a set of simulation experiments, and discuss optimal load-capacity allocation strategies.

Our results demonstrate that allocating the total free space (i.e., capacity minus initial load) to each layer proportional to the mean effective load in the layer and distributing that free space equally among the nodes within the layer ensures maximum robustness.

The proposed framework for different failure conditions allows us to analyze the two overload conditions presented and can be extended to explore more complex interdependent relationships.

MA64

Building B Level 4 B404

Supply Chain Interactions – Humans, Algorithms, Bots...

Invited Session

Behavioral Operations Management

Chair: Michael Becker-Peth, Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands

1 - Negotiating Against Chatbots: Do I Expect/Get Better Outcomes from a Human OR a Bot Counterpart? Let's Negotiate Again!

Klaus Garrido, Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands, Michael Becker-Peth, Jelle de Vries

The adoption of Artificial Intelligence (AI) in procurement is rapidly accelerating, with significant implications for operational efficiency. This research explores the operational impact of artificial counterparts in business-to-business (B2B) negotiations. Through a lab experiment with 348 participants, we investigate how the nature of negotiation counterparts (human or AI) affects negotiation outcomes, profit expectations, private information disclosure, and the preference for bots as future negotiation counterparts. Our findings reveal that 1) individuals have higher profit expectations from negotiations with bots compared to negotiations with humans. 2) Negotiators achieve comparable agreement rates but higher profits with bots, challenging prior research suggesting lower cooperation and agreement rates in human-bot interactions. Moreover, 3) exceeding expected profits significantly influences the choice between human and artificial counterparts. Specifically, having profited more against bot counterparts than other humans seems to increase the likelihood of choosing an artificial counterpart in a final negotiation. Finally, our post-hoc analysis reveals 4) over time, humans learn to exploit the predictability of the artificial counterpart by disclosing non-verifiable private information. Therefore, this study has tangible implications for practitioners aiming to implement automated negotiation systems in procurement. Managers should carefully evaluate the long-term effectiveness of artificial negotiators due to exploitation vulnerability. Thus, we raise awareness of the risks associated with deploying predictable AI negotiators that are collaborative, trusting, and honest.

2 - How to Account for Behavioral Newsvendors: the Robust Buyback Contract to Address Response Uncertainty

Michael Becker-Peth, Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands, Christina Imdahl, Kai Hoberg

Normative (expected-profit-maximizing) theory assumes that decision-makers are fully rational, but in reality, they deviate from the optimal response. In contract negotiations, contract parameters are often optimized based on the assumption of rational behavior. Deviations from rational behavior can lead to significant costs for all parties involved. To address this, we propose robust optimization to obtain contract parameters that account for deviations from rational behavior.

We apply this optimization approach to obtain the robustly optimal contract parameters for the buyback contract when there is limited knowledge of the buyers' responses. We develop a parameterization of the robust contract, called the naive robust contract, that is based on the pull-to-center effect.

We compare the naive robust contract to the normative contract by simulating orders that follow certain known behavioral models, such as mean-anchoring, bounded rationality, or a behavioral model explicitly for the buyback contract.

Additionally, we test different the normative, the naive robust and various behavioral contracts in a lab experiment with actual human decision-makers. The results demonstrate that a robust contract, based on simple assumptions, can outperform the normative contract in terms of risk while not sacrificing average supply chain profit. The robust contract also weakly dominates all other behavioral contracts, meaning that it significantly outperforms them in either average supply chain profit or risk. This approach offers a practical way to set contract parameters without relying on assumptions about the distribution of responses, except for the support.

3 - Strategic Interactions in Surplus Drug Donations: A Game Theory and Experimental Approach

James Hill, Ohio State University, Columbus, OH, United States, Telesilla Kotsi, Soh Hyun Chu

We employ game theory to model strategic interactions between Corporate Social Responsibility and Supply Chain teams in pharmaceutical firms, focusing on surplus drug donations. Controlled experiments explore how monetary incentives and operational transparency influence supply chain managers' decisions, identifying factors that align donation strategies with operational objectives in complex environments.

4 - The Effect of Cognitive Load on Algorithmic Reliance

Han Oh, Tilburg University, Tilburg, Netherlands, Eirini Spiliotopoulou, Rogelio Oliva

Research on behavioral decision-making has shown that flexibility and adaptability are among the strengths of human decision-makers. However, there is limited research on how to integrate these strengths into algorithmic decision-making systems. In this study, we explore how decisions are made under cognitive load and how decision support systems can help individuals better allocate their cognitive resources to improve decision quality.

5 - Is AI An Algorithm by Any Other Name? Behavioral Reactions to Model- and Ai-Based Demand Planning Algorithms

Finnegan McKinley, University of Arkansas, Fayetteville, AR, United States, Rebekah Brau, John Aloysius, Adriana Hofer

The rapid implementation of AI in demand planning offers the potential to improve forecast accuracy. However, users' behaviors toward AI advice might undermine potential gains. This study unveils how human behavior may differ due to the mere revelation of algorithm type—AI- or model-based—and algorithm signals of bias and learning. Using signaling theory, we hypothesize that algorithm type and signals of bias and learning influence user assessment of algorithm credibility, subsequently impacting forecast adjustment behavior. We test for differences in user adjustments based on algorithm type and signals using two laboratory experiments and a quasi-natural field experiment with over 1.1 million observations from a multinational retailer. We find that users adjust AI- and model-based algorithms similarly for a demand planning task in the lab. We also find that when algorithms signal positive bias, users adjust downwards, and when algorithms signal negative bias, users adjust upwards. However, we find that these upward and downward adjustments are amplified for AI-based algorithm use. Concerning learning, we find that users adjust to learning algorithms by decreasing their absolute forecast adjustments, but we do not find a difference in this adjustment behavior for algorithm type. We fully replicate our

significant results in the field, finding that experienced demand planners exhibit an amplified response behavior towards bias in AI algorithms. Our research advances AI recommendation adherence research by showing that the unique context of demand planning evokes a complex response to AI.

MA65

Building B Level 4 B405

GenAI in Platform Operations: Trust, Competition, Algorithms and Models

Invited Session

Artificial Intelligence

Chair: Ruohan Zhan, University College London, London, United Kingdom

Co-Chair: Ziyi Cao, University of Texas at Dallas, Richardson, TX, United States

1 - Exposure, Readiness, and Firm Valuation of Generative AI

Haonan Yin, University of California - Irvine, Irvine, CA, United States, Sanjeev Dewan, Fei Ren

Generative AI can be considered as a “General Purpose Technology” that has the potential to create broad impact across industries and economies, yet we have little understanding of how these transformative AI technologies are helping firms create and capture value. Recent research has shown that a substantial portion of the corporate workforce is exposed to GenAI and that firms with higher GenAI Exposure experience positive cumulative abnormal returns (CARs). We build upon this research and provide a more nuanced understanding of the mechanisms driving improvements in firm valuation by deconstructing exposure into three dimensions based on business functions: product, process, and customer. Additionally, we argue that Exposure alone does not guarantee added firm value, but that Exposure has to be matched with complementary AI strategies and technological investments, which we call strategic AI Readiness. In results, we found that firms with both high Exposure and high Readiness have positive and significant CARs, averaging 0.94%, over a two-day event window. In contrast, firms with high Exposure but low Readiness exhibit positive but statistically insignificant CARs, indicating limited benefits from GenAI in the absence of sufficient readiness. Interestingly, firms with low Exposure but high Readiness experience significantly negative CARs, averaging -0.32%, suggesting that misaligned AI adoption can harm firm value. These findings highlight the risks associated with premature or misaligned GenAI initiatives. Overall, this study provides actionable insights for managers, underscoring the importance of synchronizing AI strategies with organizational infrastructure and workforce capabilities to fully leverage the potential of generative AI.

2 - Career1: Reasoning Models for Career Path Prediction via Reinforcement Learning

Shuzhang Cai, The University of Texas at Dallas, Richardson, TX, United States, Ying Xie, Shaojie Tang

Modern labor markets present individuals with complex and dynamic career choices. Online platforms are increasingly use predictive algorithms to recommend jobs and anticipate mobility patterns. The progress in machine learning makes it possible to analyze large-scale career data and uncover patterns of mobility. The research on career path prediction aims to anticipate future job transitions by leveraging historical career trajectories and associated information. Early studies typically rely on structured information such as job titles, durations, employers, and skills, but do not take into account the free-text descriptions in resumes. Yet these descriptions contain rich information and can substantially enhance prediction. Recent advances in reinforcement learning and large reasoning models (LLMs) have opened new opportunities for problem-solving or decision-making tasks. In particular, for career data containing natural language descriptions of work experiences, LLMs can leverage their strong capacity for text processing and understanding to extract richer semantic information and thereby enhance the accuracy of prediction. In this paper, we employ a base LLM with the reinforcement learning method Group Relative Policy Optimization (GRPO) to tackle this problem. We propose CAREER1, a set of novel reasoning models through the integration of knowledge distillation supervised fine-tuning and reinforcement learning. Specifically, we modify the reward function in GRPO, which results in two CAREER1 variants optimized under different learning signals. Experiments on three real-world career path datasets demonstrate that CAREER1 variants improve the average Recall @5 by 2.3%, Recall@10 by 2.6% and MRR by 0.023 relative to the state-of-the-art baseline models.

3 - Past Tells, Future Sells: Automatic Fashion Product Creation with Generative AI

Jianghao Lin, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Ruohan Zhan, Zi Ling

E-commerce has revolutionized retail, yet its traditional workflows remain inefficient, with significant time and resource costs tied to product design and manufacturing inventory. We introduce a novel system deployed at Alibaba that leverages AI-generated fashion products to address these challenges with personalized text-to-image generation. In this way, we enable an innovative business model called "sell it before you make it", where merchants can design fashion items and generate photorealistic images with digital models based on textual descriptions. Only when the items have received a certain number of orders, do the merchants start to produce them, which largely reduces reliance on physical prototypes and thus accelerates time to market. This paradigm has been fully deployed and tested at Alibaba. The AI-generated fashion products have achieved over 13% relative improvements for both click-through rate and conversion rate compared to their human-designed counterparts, validating the revolutionary potential of AI-driven automatic fashion product creation for e-commercial platforms.

4 - Large Language Models and Large-Scale Causal Decision-Making

Shichao Han, Tencent, Shanghai, China, People's Republic of

Randomized controlled experiments (RCTs) are a scalable and reliable approach for large-scale decision-making, particularly when combined with observational data and appropriate statistical tools. The advent of large language model (LLM) agents has further improved the efficiency of causal discovery, leveraging their advanced capabilities to mimic human behavior and process unstructured data. However, before LLMs can be widely adopted for decision-making at scale, it is essential to establish and quantify the credibility of the insights they generate. By integrating causal inference techniques with artificial intelligence, platforms can enhance their ability to perform causal discovery more effectively. Additionally, we illustrate how LLMs serve as complementary tools to RCTs and other statistical methods, drawing on examples from Weixin.

5 - ACS: An Interactive Framework for Conformal Selection

Zhimei Ren, University of Pennsylvania, Philadelphia, PA, United States, Yu Gui, Ying Jin, Yash Nair

In this talk, I will introduce adaptive conformal selection (ACS), an interactive framework for model-free selection with guaranteed error control. ACS extends Jin and Candès (2023b) to allow for adaptive data analysis, where decisions can be made on the fly based on partially revealed data. The key to ACS is a carefully designed screening principle that reveals data, which allows us to update the selection rule while maintaining rigorous control of the false discovery rate (FDR). Under the general adaptive framework, we provide concrete instantiations of ACS for power improvement, diversified selection, and incorporating new labeled data. We demonstrate the effectiveness of ACS through extensive numerical simulations and real data applications in large language model (LLM) deployment and drug discovery.

MA66

Building B Level 4 B406

Teaching Tomorrow's Leaders: Embedding AI and Analytics in Supply Chain Education

Invited Session

INFORMS Committee on Teaching and Learning

Chair: Ramakrishna Garine, university of north texas, Peoria, 61615, United States

Co-Chair: Senthilkumar Thiyagarajan, Medline Industries, LP, Wheeling, IL, 60090, United States

Co-Chair: Nixalkumar Patel, LG Electronics, 19 Lincoln Ave, Lake Hiawatha, NJ, 07034, United States

1 - From AI Hype to Applied Learning: Equipping Supply Chain Professionals with Real-World Ai Skills

James Gaffney, Georgia Tech Supply Chain Institute, Atlanta, GA, United States

The future of supply chain belongs to professionals who can critically assess, deploy, and lead with AI—not just talk about it. In this session, I'll share insights from Georgia Tech's Generative AI in Supply Chain course, designed for industry professionals navigating both the opportunities and risks of AI adoption. Drawing on our "AI and Analytics Ladder" framework and the article "Skills to Survive and Thrive in the Age of AI," this presentation highlights how we move beyond buzzwords to build capability. I'll cover how we use real-world use cases to demystify Gen AI tools like ChatGPT, Power BI, and Python. I'll also share our interdisciplinary teaching approach, which includes applied exercises, AI policy discussions, and critical thinking "workouts" to ensure learners don't lose the human edge. Finally, I will share our approach to keep the content current as it needs a near complete re-write for every session. Participants will walk away with example lesson plans, a curated list of hands-on tools, and a model for designing curriculum that keeps pace with the evolving demands of the AI-powered supply chain—while still rooted in supply chain fundamentals.

2 - Teaching AI-Driven Cross-Platform Data Intelligence: Embedding Automated Validation Systems in Supply Chain Education**Nihari Paladugu, Southern New Hampshire University, Columbus, OH, United States**

Supply chain disruptions cost global businesses \$2.3 trillion annually, yet 73% of graduate interns entering supply chain roles lack essential analytical skills for data governance and cross-system validation. This educational research presents the InsightOps Educational Framework, a novel pedagogy that adapts financial services' 99.99% data accuracy methodologies to supply chain management education.

The framework teaches graduate interns five core competencies: (1) attribute-level usage intelligence for data value optimization, (2) behavioral telemetry analytics for user pattern recognition, (3) AI-powered PII detection for compliance automation, (4) machine learning-based cross-platform validation across Teradata, Snowflake, and ERP systems, and (5) self-learning governance engines for policy enforcement.

Implementation with 12 graduate interns at a Fortune 500 manufacturer demonstrated remarkable outcomes: 99.2% data accuracy across platforms, 95% reduction in migration issues, and 23% salary premiums for program graduates. The methodology transforms traditional supply chain education by embedding automated validation systems thinking derived from regulated financial services.

This human-AI collaborative approach prepares graduate interns as AI orchestrators rather than automation victims, teaching systematic data governance excellence while maintaining ethical oversight. The framework offers scalable educational innovation for institutions seeking to bridge the critical skills gap in supply chain analytics leadership.

3 - Agent-Driven Decision Support in Inventory Replenishment and Exception Resolution**Harshavardhan Nannur, Walmart, Dallas, TX, United States**

Modern retail supply chains face increasing complexity, with dynamic demand, fragmented data sources, and frequent operational exceptions. This talk introduces a practical framework for integrating agentic AI into inventory replenishment and exception management workflows. We present a modular, human-in-the-loop system where generative agents monitor replenishment signals, detect anomalies, and suggest context-aware resolutions across clubs and fulfillment centers. The approach blends rule-based monitoring with LLM-powered reasoning to improve issue triage, reduce manual effort, and drive faster, more consistent decision-making. Results from real-world pilots demonstrate measurable improvements in response time and replenishment accuracy.

4 - Building Ethical AI Literacy in Future Supply Chain Professionals**Anand Chinnakannan, Quantic School of Business & Technology, Bentonville, AR, United States**

As artificial intelligence becomes central to decision-making across the supply chain—from forecasting and procurement to logistics and customer fulfillment—the need for ethical and responsible AI use is more urgent than ever. Yet, while technical education in AI and analytics is growing rapidly, ethical considerations often remain overlooked.

This session presents a framework for integrating ethical AI literacy into supply chain education by mapping the ethical risks and responsibilities at each stage of the supply chain. Key topics include bias in demand forecasting, fairness in supplier selection algorithms, labor impacts of automation, data privacy in logistics, ethical price personalization, and sustainability in reverse logistics.

Through real-world case examples and curriculum design strategies, the session equips educators and professionals with tools to prepare students not only as skilled AI practitioners but as thoughtful, responsible decision-makers. Attendees will gain actionable insights for embedding ethics into both technical and managerial training in supply chain programs.

5 - Agentic AI in Supply Chain Education: Teaching Beyond Chatbots**Shiva Kumar, Albertsons, Plano, TX, United States**

The integration of Artificial Intelligence into supply chain operations has rapidly evolved beyond traditional machine learning models and static automation tools. At the forefront of this transformation is Agentic AI, a new generation of intelligent systems capable of autonomous planning, contextual reasoning, proactive decision-making, and human-like interaction. These systems go far beyond reactive chatbot interfaces and are increasingly being used in real-world supply chain environments to manage inventory, forecast demand, optimize logistics, and coordinate supplier networks.

6 - Empowering Supply Chain Education with Real-World AI and Open-Source Optimization Tools**Herin Savla, ABF Freight, Fort Smith, AR, United States**

As the logistics and supply chain industry continues to evolve, the integration of AI and analytics has become a cornerstone of operational efficiency. However, there remains a significant gap between academic training and industry expectations—particularly in the practical application of optimization techniques. This presentation will explore how open-source tools, such as Google OR-Tools, can be effectively incorporated into supply chain and operations research curricula to bridge this gap. Drawing from my professional experience in leading large-scale analytics initiatives at ABF Freight, I will highlight the value of hands-on exposure to real-world problem-solving using accessible, industry-relevant technologies. Attendees will gain insights into how these tools can be used to model and solve complex optimization problems like vehicle routing problems, preparing students to contribute meaningfully from day one in their careers. By fostering a curriculum that emphasizes both theoretical foundations and practical applications, we can better equip the next generation of supply chain professionals to meet the demands of a data-driven world.

7 - Developing AI and Analytical Skills in Supply Chain Professionals to Build Lean, Agile, Resilient, and Sustainable (Lars) Supply Chains**Senthilkumar Thiagarajan, Medline Industries, LP, Wheeling, IL, United States**

The supply chain profession is undergoing a profound transformation driven by rapid technological advancements, disruptions, and rising operational complexities. Traditional skill sets are no longer sufficient to drive meaningful impact in supply chains. To thrive in today's dynamic environment, supply chain professionals must develop capabilities in AI, machine learning, and advanced analytics to harness the power of Industry 4.0 technologies – including digital twins and the Internet of Things (IoT). These technologies are critical enablers of lean, agile, resilient, and sustainable supply chains, and they also generate massive volumes of data to support informed, insight-driven decision-making in supply chain operations.

This presentation explores how academic institutions and organizations can prepare current and future supply chain leaders to meet these emerging demands. It outlines a structured approach for **teaching, evaluating, and recognizing AI and analytics competencies**, drawing from both supply chain management and learning literature. As organizations move toward digital, resilient and sustainable supply chains, developing human capability becomes just as important as technological investment. Special attention is given to methods for upskilling early-career professionals, mid-level managers, and supply chain practitioners transitioning into digitally intensive roles. By linking technological adoption with human capability development, this session aims to close the gap between research and real-world application, innovation and execution, ensuring that the workforce is equipped to lead the next era of supply chain transformation.

MA67

Building B Level 4 B407

Recent Advances in Optimization Under Uncertainty

Invited Session

Data Mining

Chair: Muyun Lu, University of Houston, Houston, TX, United States

Co-Chair: JUN ZHOU, University of Houston, Houston, TX, 77054, United States

Co-Chair: Ying Lin, University of Houston, Houston, TX, United States

1 - A Decomposition Method for Integrated Chance-Constrained Stochastic Programming with Application to Optimal Vaccine Allocation**Xin-Yu Zhuang, Texas A&M University, College Station, TX, United States, Lewis Ntaimo**

Vaccination is crucial for controlling epidemics especially for viruses that mutate or have different strains with uncertain transmission characteristics. To prepare for future outbreaks, data-driven models for optimizing vaccination strategies are essential. Integrated chance-constrained stochastic programming (ICC-SP) has been proposed for this purpose, but solving large-scale instances of ICC-SP models efficiently is challenging. To address this, we propose a decomposition cutting-plane method for ICC-SP model for optimal vaccine allocation. This ICC-SP model is a multi-community, household-based epidemiological model that incorporates census demographic data, vaccination status, age-related differences in disease susceptibility and infectivity, and vaccine efficacy. We report on computational study based on real data for flu for the State of Texas and for COVID-19 for the South Africa setting.

2 - Optimal Mechanism Design with Constraint-Aware Bayesian Optimization**Muyun Lu, University of Houston, Houston, TX, United States**

Mechanism optimization seeks to identify an optimal mechanism that maximizes a designer's objective while satisfying equilibrium constraints, such as Nash Equilibrium. This task is computationally challenging, often formulated as a Mixed-Integer Nonlinear Programming problem, which is NP-hard and lacks efficient general-purpose solvers. To address this challenge in the context of the sharing economy, exemplified by peer-to-peer ride-sharing, we develop a nonlinear optimization model based on the Affine Maximizer framework. We propose a novel algorithm, Optimal Mechanism Design with Constraint-aware Bayesian Optimization (MDCBO), to solve this model. MDCBO leverages constraint-aware Bayesian optimization to intelligently search the design space while adhering to constraints, as the Affine maximizer determines the allocation rule and provides crucial feedback to meet equilibrium requirements. This synergistic approach enables the efficient discovery of high-performing mechanisms in computationally demanding environments. Theoretical analysis and numerical studies are conducted to demonstrate the solution quality and computational efficiency of MDCBO for the optimal mechanism design.

3 - Robust Remanufacturing Planning with Interval Matrix-Based Uncertainty Set**JUN ZHOU, University of Houston, Houston, TX, United States**

We study the problem of remanufacturing planning under uncertainty arising from statistical estimation errors. A critical aspect of determining optimal remanufacturing timing is accurately modeling the system's state transition dynamics. However, such estimations often suffer from inaccuracy due to limited or noisy data, leading to significant performance degradation when conventional methods applied. To mitigate the impact of uncertainty, we formulate the problem as a robust Markov decision process (RMDP), where the system's true state transition probability matrix is unknown but assumed to lie within an ambiguity set that captures all possible realizations consistent with observed data. We focus on ambiguity sets defined by interval matrices, which offer a tractable and interpretable means of representing uncertainty in transition dynamics. A tailored solution methodology that efficiently computes optimal robust policies is developed. We further establish structural properties of optimal robust policies and provide theoretical insights into the interval-matrix-based ambiguity sets. A computational study on the NASA turbofan engine shows that our data-driven decision framework consistently yields better worst-case performances and higher reliability of the performance guarantee.

4 - Distributionally Robust Universal Classification: Bypassing the Curse of Dimensionality**Siyuan Chen, Georgia Institute of Technology, Atlanta, GA, United States, Weijun Xie**

The Universal Classification (UC) problem seeks an optimal classifier from a universal policy space to minimize the zero-one loss. However, the conventional empirical risk minimization often leads to overfitting and poor out-of-sample performance. To address this limitation, we introduce the Distributionally Robust Universal Classification (DRUC) formulation, which incorporates distributional robustness via the Wasserstein distance centered at the empirical distribution. To manage the infinite-dimensional nature of the DRUC policy space, we develop its in-sample DRUC counterpart, which allows for a more tractable reformulation while preserving robustness properties. We prove that, asymptotically, the in-sample DRUC formulation converges to the original UC formulation and is equivalent to the DRUC formulation. Under mild conditions, we provide non-asymptotic finite-sample performance guarantees. Furthermore, we derive a mixed-integer linear programming (MILP) reformulation to obtain the optimal in-sample DRUC policy and propose an efficient 2-approximation algorithm. Our numerical experiments show the efficiency of the proposed approximation algorithm and demonstrate the superior out-of-sample performance of the in-sample DRUC formulation.

5 - Performance of Myopic Replenishment Policies for Retail Inventory Management: a Distributionally Robust Approach**Mohammad Razzaghi Kashani, University of Houston, Houston, TX, United States, Mehdi Farahani**

In this paper, we study the problem of determining optimal replenishment quantities for multiple products over multiple periods in retail stores subject to limited shelf space and minimum presentation constraints. Our goal is to maximize total profit by accounting for sales revenue and procurement costs in environments characterized by uncertain product demand.

We propose a myopic replenishment policy embedded within a distributionally robust optimization (DRO) framework. The DRO approach enables us to make minimal assumptions about demand distributions, relying solely on their mean and variance, thereby reducing forecasting complexity and improving practical applicability. Using dynamic programming, we analytically demonstrate the optimality of myopic policies, particularly when store capacity is significantly constrained relative to potential inventory levels—a frequent scenario faced by retailers due to SKU proliferation and rapid product turnover. Our theoretical analysis provides structural insights into optimal inventory strategies by identifying explicit conditions under which myopic policies achieve optimal performance.

We validate our theoretical results through extensive numerical experiments. The experiments confirm that the proposed myopic policy achieves excellent performance across diverse operational scenarios. Furthermore, the simplicity, computational efficiency, and limited data requirements of the policy significantly enhance its managerial appeal, facilitating straightforward implementation in practice.

Our findings underscore a critical insight for retail operations, highlighting that although sophisticated inventory models may theoretically promise superior outcomes, real-world constraints—particularly the limited shelf space frequently faced by retailers in practice—often make simpler, intuitive replenishment strategies more effective and better aligned with actual business needs.

MA68

Building B Level 4 B408

Variable Importance

Invited Session

Data Mining

Chair: Cynthia Rudin, Duke University, Durham, NC, United States

Co-Chair: Srikar Katta, Duke University, Durham, NC, United States

1 - Variable Importance: a Critical Overview

Emanuele Borgonovo, Bocconi University, Milan, Italy, Elmar Plischke, Xuefei Lu, Cynthia Rudin

Variable importance measures are a central tool in explainable AI. They help analysts assess how much each input feature contributes to a model's output. However, the plethora of available methods has led to confusion. Not all variable importance tools serve the same purpose: some provide indications for individual predictions, while others yield a global importance index at the dataset level. Also, some methods are model-specific while others are model-agnostic. It is then essential that analysts define the variable importance question up front and do not rely on the same hammer in all situations.

Also, when calculating feature importance measures issue might emerge if the corresponding algorithm forces the model to extrapolate. In this presentation, we review these and several other critical aspects that emerge in the use of variable importance measures.

2 - Variable Importance on Steroids

Cynthia Rudin, Duke University, Durham, NC, United States

I will overview modern work on model-free variable importance. I will introduce model class reliance, variable importance clouds and diagrams, concept whitening, and put some of the other work in this session into perspective, including the Rashomon Importance Distribution.

3 - A Rashomon Approach to Variable Importance with Distribution Shift

Jon Donnelly, Duke University, Durham, NC, United States, Srikar Katta, Emanuele Borgonovo, Cynthia Rudin

Variable importance analysis is an essential tool for knowledge discovery, particularly in settings where controlled trials are prohibitively expensive or impractical. Through variable importance, we can suggest factors that might be important in determining an outcome using only observational data.

However, existing methods for variable importance do not account for the impact of *unobserved* variables on their analysis. Moreover, existing methods tend to measure importance with respect to a single model, over a single, finite dataset. These shortcomings can lead to misleading results.

We address each of these problems through a Rashomon approach to variable importance under distribution shift (RAVIDS). We develop a theoretical connection between the impact of unobserved variables and the Rashomon set -- the set of all near optimal models -- and provide finite sample bounds on the estimation of the Rashomon set. We show through theory and simulation studies that RAVIDS captures the true importance of variables to the true data generating process with high probability, even when there are important unobserved variables.

4 - Targeted Learning for Variable Importance

Xiaohan Wang, Cornell University, Ithaca, NY, United States, Yunzhe Zhou, Giles Hooker

Variable importance is one of the most widely used measures for interpreting machine learning with significant interest from both statistics and machine learning communities. Recently, increasing attention has been directed toward uncertainty quantification in these metrics. Current approaches largely rely on one-step procedures, which, while asymptotically efficient, can present higher sensitivity and instability in finite sample settings. To address these limitations, we propose a novel method by employing the targeted learning (TL) framework, designed to enhance robustness in inference for variable importance metrics. Our approach is particularly suited for conditional permutation variable importance. We show that it (i) retains the asymptotic efficiency of traditional methods, (ii) maintains comparable computational complexity, and (iii) delivers improved accuracy, especially in finite sample contexts. We further support these findings with numerical experiments that illustrate the practical advantages of our method and validate the theoretical results.

MA69

Building B Level 4 B409

Human Behavior in the AI Era

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Zhaohui Jiang, Carnegie Mellon University, Pittsburgh, PA, United States

Co-Chair: Clare Snyder, NYU Stern, New York, NY, United States

1 - Unlocking the Value of Real-Time AI Assistance: Who Benefits, and Why?

Zhaohui Jiang, Carnegie Mellon University, Pittsburgh, PA, United States, Yao Cui, Qi Li

As companies increasingly integrate artificial intelligence (AI) into their services, it is becoming crucial to understand AI's impact on service users. One novel application of AI is to provide real-time assistance throughout the service process, featuring a more dynamic and interactive human-AI relationship. In this paper, we investigate how human users derive value from such real-time AI assistance and how this value varies according to users' experience levels. We address these questions through a collaboration with an international car-sharing platform that recently underwent a technology upgrade involving an AI-powered driver monitoring and feedback system. Combining theoretical insights with empirical analyses, we uncover a heterogeneous effect characterized by a U-shaped relationship: users with low or high experience levels benefit from AI assistance, while users with medium experience levels do not. Our findings further reveal that different user groups derive value from AI in distinct ways: low-experience users derive value passively due to their reliance on AI assistance, whereas high-experience users actively enhance AI's value by adapting their behaviors to mitigate the costs associated with AI during the real-time interaction. On the other hand, medium-experience users struggle to overcome the cost-benefit trade-off of AI, as they derive less benefit from AI's assistance compared to their less experienced peers and lack the adaptability of their more experienced peers to mitigate its costs. These insights are useful for guiding AI adoption and algorithm customization in service applications, and provide broader implications for other human-AI interaction processes, such as AI-assisted medical diagnosis and ChatGPT.

2 - Removing Phones from Classrooms Improve Educational Outcomes

Alp Sungu, Wharton, Philadelphia, PA, United States

We conducted a randomized controlled trial with 17,600 students to assess the causal effects of smartphone bans in classrooms. Implementing mandatory phone depositing to dedicated boxes during lectures improved student grades, particularly among lower-performing, first-year, and non-STEM students, with an average GPA increase of approximately 0.078 standard deviations. Importantly, students exposed to the ban became noticeably more supportive of phone-use restrictions, perceiving greater benefits from these policies and displaying reduced preferences for unrestricted access. Such enhanced student receptivity to restrictive digital policies, could reinforce the policy effectiveness through firsthand experience. Despite a mild rise in reported fear-of-missing-out, there were no significant negative effects on over-all student well-being, academic motivation, or digital use. Random classroom spot checks indicated fewer instances of chit-chat and disruptive behaviors among students, along with reduced phone usage and increased engagement among teachers in phone-ban classrooms—suggesting a classroom environment more conducive to learning. Spot checks also revealed that students appear more distracted, possibly due to withdrawal from habitual phone checking, yet, students did not report being more distracted. These findings suggest that classroom phone bans pose a low-cost, modest but “quick-win” policy recommendation to improve educational outcomes—especially for academically vulnerable groups.

3 - Guardrails on Biases? the Impact of Limiting Human Adjustments on Inventory Performance

Yixin Iris Wang, Gies College of Business, University of Illinois Urbana-Champaign, Champaign, IL, United States, Minyu Ye, Jun Li, Stephen Leider

In practice, managers often deviate from the ordering quantities recommended by inventory algorithms, citing private information not captured by the system. This study leverages a field experiment in offline pharmacy stores to examine the impact of restricting managers' manual adjustments on store performance. We further investigate the underlying drivers of these adjustments to determine whether the resulting performance changes are driven by a reduction in behavioral biases, the effective use of private information, or both.

4 - Too Neat To be True? Bias and Consumer Response with AI-Generated Review Summaries

YUNZHOU ZHONG, University of Pennsylvania, Philadelphia, PA, United States, Simin Li, Jiding Zhang

Review summaries generated by large language models (LLMs) are increasingly featured on user-generated content (UGC) platforms. Using the generative AI-generated hotel summaries and human-written reviews on TripAdvisor, this early-stage study investigates potential biases in AI-generated summaries of hotel reviews—specifically, how these summaries vary with characteristics of the underlying review pool, such as size and rating dispersion. We aim to understand whether AI summaries fairly represent the original content and how any summarization bias may shape consumers' expectations of service quality. We also examine how discrepancies between expectation and actual experience influence subsequent customer reviews. Additionally, we measure the concentration of demand across hotels, and explore the effect of AI-generated review summaries on market structures. By leveraging longitudinal data from TripAdvisor, our analysis aims to uncover the behavioral and market implications of integrating generative AI into review platforms.

5 - Green Intervention and User Compliance: a Field Experiment

Yi Song, Tsinghua University, Beijing, China, People's Republic of, Yao Cui, Da Zhang

Smart automated control of energy use is an emerging approach to resource conservation. However, its effectiveness can be limited by insufficient user compliance. In a randomized controlled trial conducted across public office buildings in China, we examine the impact of introducing a smart A/C emissions control system on electricity consumption under different control strategies. Our findings suggest that simply increasing the target temperature may not yield significant energy savings, but control strategies that adjust target temperatures more frequently or dynamically in response to user behavior can be more effective. These results underscore the importance of considering user compliance in the implementation of green technology and offer practical solutions for enhancing energy-saving outcomes.

Monday, October 27, 9:45 AM - 10:45 AM

Building A Level 4 Auditorium

Plenary - Ramayaa Krishnan

Plenary Keynote

Plenary

1 - Two Heads are Better than One Redux: AI, OR/MS and Society

Ramayya Krishnan, Carnegie Mellon University, Pittsburgh, PA, United States

In 1986, Nobel Laureate Herbert Simon delivered the TIMS/ORSA Joint National Meeting plenary titled “Two Heads Are Better Than One: The Collaboration between AI and OR”. Close to 40 years later, we are seeing the rapid development of AI technologies opening up dramatic new opportunities for synergistic collaboration between AI and OR/MS. From large language models to large reasoning models, from Gen AI for text to multimodal AI, and from chatbots to agentic AI, these technologies are important additions to our decision analytic and technology toolkits. However, adoption has been lagging due to a significant gap in understanding how to evaluate these technology artifacts and guide their deployment into workflows and systems to support consequential decision making. In this plenary talk, I posit the need for two types of synergistic collaboration to advance the science of measurement and evaluation required to achieve this objective. The first synergy opportunity – OR for AI (ORFAI) – is required to enable dynamic, multi-criteria analysis and evaluation of AI artifacts (e.g., LLM models, agents) as “systems”. The second – OR/AI for Societal Systems (ORAIS) – harnesses AI and OR/MS in consequential scenarios and workflows, and evaluates the implications of such deployment for workers and their reskilling/upskilling. I will discuss the need for robust evaluation infrastructure that will leverage this desired synergy between OR/MS and AI, and the significant implications for education, business and public policy.

Monday, October 27, 11:00 AM - 12:15 PM

MB01

Building A Level 3 A301

Gurobi Optimization/Responsive.net

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Building Trust in Optimization

Caroline Weinberg, Gurobi Optimization, Boston, MA, United States

In this session we will explore some of Gurobi's advanced features through the lens of solution explainability, including solution pools, infeasibility analysis, and multi-scenario analysis. Explainability in practice requires both technical skills and successful communication with stakeholders who don't know how Gurobi works (and that's ok!). We show how these tools are not just for modeling – they can help users interpret optimization results and build trust their solutions.

2 - Using An Online Factory Management Simulation to Teach Operations Management

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

Littlefield for Operations is a competitive online simulation of either a factory or a medical laboratory that has been used by more than half a million students in 500+ universities in 60+ countries over the past 26 years. In this session, we will discuss how professors use the newest version of Littlefield for Operations (2.0) to effectively teach foundational operations management topics such as process analysis, capacity planning and inventory control to students at various levels in various course formats.

MB02

Building A Level 3 A302

Embracing GenAI: Navigating the New Landscape of Knowledge Work

Panel Session

Committee's Choice

Co-Chair: Maryam Alavi, Georgia Tech, Atlanta, GA, United States

1 - Moderator Panelist

Maryam Alavi, Georgia Tech, Atlanta, GA, United States

Over the past decades, enterprises have successfully harnessed advances in computer, communication, and robotic technologies to automate routine tasks at scale, such as payroll management, inventory control, and manufacturing processes. This automation has delivered substantial efficiency gains, cost reductions, and productivity

enhancements across diverse sectors. Despite these successes, professional knowledge work has largely remained resistant to automation due to technological constraints. The advent and rapid evolution of generative AI (GenAI), however, are poised to dramatically alter this landscape. GenAI tools now offer unprecedented, human-like capabilities, including idea generation, natural language processing, software development, and multimedia content creation (text, image, video).

These capabilities present significant opportunities for organizations to transform knowledge-intensive tasks previously considered beyond the reach of traditional automation. Integrating GenAI into knowledge work promises substantial benefits, such as accelerated innovation, heightened productivity, and enhanced competitive advantage. Yet, deploying GenAI at scale involves navigating complex, multi-dimensional challenges that span technological infrastructure, individual behaviors, and organizational culture and practices. This panel will explore, and clarify GenAI's transformative potential for knowledge work, addressing crucial issues and insights from technological, behavioral, and organizational perspectives, and emphasizing the critical need for dynamic, continuous alignment among these domains.

2 - Panelist

Warren Hearnese, OptiML AI, Atlanta, GA, United States

3 - Panelist

Deepa Krishnan, IBM Americas, Atlanta, GA, United States

4 - Panelist

Reza Mousavi, University of Virginia, Charlottesville, VA, United States

MB03

Building A Level 3 A303

Toward Intelligent Multi-Criteria Decision-Making: Methods and Applications

Invited Session

Multi Criteria Decision Making

Chair: Birsan Karpak, Youngstown State University, Canfield, OH, United States

Co-Chair: Ilker Topcu, Istanbul Technical University, Istanbul, Turkey

1 - Exploring the Interaction of Pressure Sources and Types on Sscm Practices: An ANP Model Approach

Birsan Karpak, Youngstown State University, Canfield, OH, United States, Alina Marculetiu, Cigdem Ataseven, Ilker Topcu

The growing emphasis on sustainability has intensified external and internal pressures on firms to adopt Sustainable Supply Chain Management (SSCM) practices. These pressures—exerted by entities such as governments, customers, suppliers, and employees—manifest through various mechanisms, including coercive (regulatory), normative (social), and mimetic (competitive) influences. While these individual factors are well-documented, their relative importance and interdependencies remain underexplored. This study addresses that gap by proposing an Analytic Network Process (ANP) model to evaluate and prioritize the interactions between pressure sources and pressure types, providing a more comprehensive understanding of how they collectively influence SSCM practices across internal, upstream, and downstream operations.

Drawing on a synthesis of 93 articles published between 1997 and 2022, the research identifies key factors and develops a network-based framework that captures complex, multi-directional relationships. Expert input is collected through pairwise comparisons involving regulators, corporate sustainability leaders, NGOs, and industry analysts. The ANP model then quantifies the influence of each source-type combination, revealing which pressures most effectively drive SSCM adoption.

Initial findings indicate that coercive pressure from regulators and normative pressure from investors are among the most impactful. Additionally, the model highlights how certain interactions amplify or mitigate the effectiveness of others, offering actionable insights for firms and policymakers. By moving beyond linear cause-effect assumptions, this study contributes to SSCM literature with a nuanced, systems-based approach.

Ultimately, the findings aim to inform strategic, context-sensitive interventions that enhance supply chain sustainability and resilience in an increasingly complex global environment.

2 - An Evaluation Framework for Blockchain Selection in Supply Chain

Deepa Iyer, Cleveland State University, Cleveland, OH, United States, Birsan Karpak, Ilker Topcu

Selecting the appropriate blockchain platform is a critical yet complex decision in supply chain management, given the evolving technological landscape and the absence of robust, adaptable evaluation frameworks. This study presents a comprehensive decision-making model using the Analytic Network Process (ANP) to guide platform selection by capturing the interdependencies among blockchain adoption enablers and evaluating the impact of alternative blockchain technologies. Drawing on an extensive literature review and expert input from industry and academia, key enablers—ranging from technical infrastructure to regulatory environment—are identified and structured into interrelated clusters. The ANP model incorporates these factors alongside leading platforms such as Ethereum Quorum and Hyperledger Fabric, allowing for a holistic comparison that reflects real-world complexity. Pairwise comparisons are conducted to derive priorities, and a Limit Matrix is computed using SuperDecisions software to synthesize results. Findings indicate that platform influence on enablers is often overlooked in existing models, limiting their applicability. Our framework addresses this gap, highlighting how different blockchain platforms affect criteria such as scalability, transparency, and integration potential. The study also considers dynamic conditions through sensitivity analysis, supporting more resilient and adaptable decision-making. Additionally, this research identifies emerging trends—particularly the growing role of artificial intelligence in blockchain integration and sustainability-focused developments, such as proof-of-stake consensus mechanisms. These insights pave the way for future studies to explore AI-enhanced blockchain strategies in supply chains. This work presents a novel, practical tool for decision-makers seeking to enhance supply chain transparency, efficiency, and resilience by selecting a tailored blockchain platform.

3 - Learning-Based Multi-Criteria Decision-Making Model for Site Selection Problems

Mahid Ahmed, University of Southern Mississippi, Hattiesburg, MS, United States, Ali Dogru, Chaoyang Zhang, Chao Meng

Choosing the right site for industrial facilities is essential for building efficient, profitable, and sustainable supply chains. Yet, finding the best location is no easy task. It involves navigating a mix of logistical, socio-economic, environmental, and climate-related factors. Traditional methods like Multi-Criteria Decision Making (MCDM) often depend on expert judgment, which can introduce bias, while exact optimization models focus heavily on proximity and struggle with computational limits as the number of site options grows. To tackle these challenges, we introduce a data-driven framework called the Learning-Based Multi-Criteria Decision Making (LB-MCDM) model. This approach blends machine learning, MCDM, and GIS-based spatial analysis to assess the suitability of potential sites in a more objective and scalable way. The model uses GIS raster analysis to score and rank candidate locations based on weights learned from ML algorithms. We apply this framework to a real-world case study in Mississippi, evaluating sawmill locations using nine key features: national land cover, distances to roads, rail lines, and urban areas, unemployment rate, terrain slope, precipitation, market revenue, and timber supply and demand. The model's predictions show strong alignment with where sawmills are currently located, validating its practical relevance. Beyond offering strategic insights, LB-MCDM can also assist traditional optimization models by reducing the number of candidate sites, making it easier to solve complex, NP-hard facility location problems.

4 - Decision Support System with a Multicriteria Approach for Prioritizing Credit Cards by Customers

VINICIUS TOMAZ, Technological Institute of Aeronautics, São Paulo, Brazil, ARSENIO GOMES NETO, Jonatas Almeida

The credit card market has become increasingly competitive in recent years in many countries and now offers a wide range of options to consumers. From this perspective, individuals trying to choose a credit card may find it difficult to compare the various perks and fees companies offer them. To assist this large group of people, a Decision Support System (DSS) was developed to allow each decision maker to tailor their criteria and identify which card is best suited to their needs, considering several options, using the SMARTER method. The solution proposed is scalable to support consumers around the world in wisely comparing different credit cards, which might stimulate

competitiveness among credit card operators and possibly lead to an improvement in the benefits offered to the public each operator wants to achieve. An example of this application is presented, analyzing 10 alternatives across 5 criteria. The model's solution indicated that "Credit Card 1" would be the most appropriate recommendation for this decision maker. Following the sensitivity analysis, the solution demonstrated strong robustness: after simulating one million scenarios to understand how the uncertainty introduced by the ROC approximation could affect the problem, the top-ranked alternative maintained its position in 97.48% of the cases.

MB04

Building A Level 3 A304

Green and Sustainable Finance

Invited Session

Finance

Chair: Victor DeMiguel, London Business School, London, United Kingdom

Co-Chair: Milind Goel, London Business School, London, HA9 0QE, United Kingdom

1 - Natural Disasters and US Small Business Performance: Heterogeneous Effects

Julien Maire, Stanford University, Stanford, CA, United States, Alec Madayan, Kay Giesecke

We develop causal survival forest estimators for the impact of natural disasters on the behavior of small business borrowers in the United States. Using damage data from SHELDUS and loan-level performance data from the Small Business Administration for the period spanning 2008-2022, our Conditional Average Treatment Effects (CATE) estimates imply that the probability of charge-off increases by 8% after 18 months for the most impacted businesses, and the probability of prepayment decreases by 2% on average after 6 months for loans in affected areas. In addition, we find significant heterogeneity in treatment effects: small, short-term, non-franchise, out-of-state loans are the most vulnerable. We also find that these effects persist over time. Our results have important implications for the management of physical climate risk by policymakers and financial institutions.

2 - Physical Climate Risk for Mortgage Markets

Johannes Fuest, Stanford University, Stanford, CA, United States, Julien Maire, Kay Giesecke

We develop meta learning estimators for the causal impact of physical climate risk on the behavior of US residential mortgages and agency mortgage-backed securities (MBS). Using county-level loss data from SHELDUS, storm tracks from HURDAT2 and loan performance data from Fannie Mae and Freddie Mac, we estimate Conditional Average Treatment Effects (CATE) for loan-level transition probabilities for multiple time horizons. Our results show a significant uplift in severe delinquency probabilities for loans in zip-code areas experiencing damages greater than \$1,000 per capita, with heterogeneous effects driven by borrower credit score, debt-to-income ratio, original loan-to-value, and interest rate. We also observe heterogeneous effects on prepayment probabilities. We explore the implications of these findings for MBS markets.

3 - Portfolio Optimization for R&D Projects: a Real Options Dynamic Programming Approach

Ruixun Zhang, Peking University, Beijing, China, People's Republic of, Leonid Kogan, Andrew Lo, Qingyang Xu

We propose a dynamic strategy to optimize a portfolio of research and development (R&D) projects with correlated binary outcomes, which we call dynamic optimization of R&D options (DORADO). DORADO utilizes project correlation as a knowledge discovery mechanism in order to exercise real options with the information learned from previous outcomes to dynamically adjust future R&D plans. We focus on a class of R&D projects, known as long shots, that are difficult to manage using standard approaches. We derive the DORADO-optimal strategy and its corresponding financial value for these portfolios under batch testing, block correlation, and risk-sensitive objectives. DORADO generates an improved efficient frontier with a higher expected payoff and lower volatility than the comparable parallel development of these portfolios. Depending on investor preferences, the financial value of such an investment portfolio is maximized when the correlation between projects is either very low or very high, which reflects the tradeoff between the diversification gains at low correlations and the information value gained at high correlations. We provide an empirical example in the context of curative treatments for glioblastoma. Our results have important implications for managing highly correlated R&D projects that cannot be effectively diversified.

4 - Bitcoin Mining and Climate Damage

Steven Kou, Boston University, Boston, MA, United States, Min Dai, Shuaijie Qian, Ling Qin

We propose a tractable dynamic equilibrium model to study the impact of Bitcoin mining on climate damage. To incorporate miners' endogenous exit and entry with technology innovation, a critical modeling step is to draw on the idea of Ben Tahar, Soner, and Touzi (2010) to track the average operating costs rather than the exact operating costs, thus overcoming the difficulty of strong path-dependency incurred by the interaction among endogenous exit, entry, and technology innovation. The model can capture empirical co-movements of miners' computing power and mining revenue. The model predicts that Bitcoin mining is not economically sustainable in terms of long-term climate damage.

5 - Legislator Tweets About the Green Transition and the Returns of Green Versus Brown Stocks

Milind Goel, London Business School, London, United Kingdom

To study the impact of regulatory transition risk on firm value, we analyze stock returns around legislator tweets about climate change. Green stocks significantly outperform brown stocks in the one-to-ten-minute window around pro-transition tweets. The cumulative daily-average green-minus-brown portfolio return around pro-transition tweets is 6.9% higher than around anti-transition ones. For tweets that mention environmental regulations, the spread increases in the weeks leading to a congressional vote and disappears afterward. Our findings suggest that transition risk impacts the relative performance of green and brown stocks—at least partly—via a regulatory channel.

MB05

Building A Level 3 A305

Office of Naval Research Program Review 1

Invited Session

ONR

Chair: David Phillips, Office of Naval Research, Arlington, VA, United States

1 - Algebraic, Geometric, and Topological Techniques for Optimization Problems from Machine Learning

Andrew Blumberg, Columbia University, New York, NY, United States

In this project we design and study machine learning methods for problems with rich geometric structure. For example, we consider methods that incorporate the symmetries from physical laws, optimal transport techniques for cell dynamics, and time series of random geometric objects. We apply these ideas to a wide range of problems, from fluid dynamics to genomic data sets.

2 - Topological Data Analysis and Optimization

John Carlsson, University of Southern California, Los Angeles, CA, United States

We present a framework that uses topological data analysis (TDA) to enhance optimization algorithms by identifying and exploiting coarse geometric features in solution landscapes. Our approach demonstrates how persistent homology can detect meaningful patterns in optimization problems across three key areas: (1) improving local search heuristics for discrete problems like the traveling salesman problem by using TDA to identify promising solution features that persist across multiple local optima,

(2) developing efficient dual quadratic programming methods for computing alpha complexes in high dimensions without requiring full Delaunay triangulation, and (3) applying filtered simplicial complexes to computer vision problems like image correspondence to classify and avoid problematic local minima caused by repeated patterns. By focusing on topological invariants, our methods provide robust guidance for optimization algorithms, helping them navigate complex solution spaces more effectively and avoid getting trapped in poor local optima that arise from geometric artifacts in the problem structure.

3 - Data-Aware Guarantees for Mathematical Optimization and Learning

Mojtaba Soltanalian, University of Illinois Chicago, Chicago, IL, United States

As AI and optimization systems become increasingly central to high-stakes decision-making, a pressing question arises: *How can we know if a model is truly doing its job?* This talk introduces MERIT—a framework that reimagines optimization by iteratively solving a sequence of well-posed surrogate problems, each bringing us measurably closer to the true solution. MERIT offers not just solutions, but certifiable progress toward global optimality.

We begin with MERIT's impact on sparse signal recovery, where we move beyond classical regularization-based approaches. By embracing judicious overparameterization, we derive data-informed reliability metrics. These results offer a fresh lens on sparsity-driven inference; especially under noisy or uncertain conditions.

We then pivot to deep learning, where MERIT enables real-time certification of training progress. We particularly focus on the Yeganegi–Eamaz–Soltanalian (YES) training bounds: a principled family of training guarantees that provide a simple answer to a fundamental question—*Has this network been properly trained by the data?* YES bounds are data-aware, structure-aware, and offer actionable insights into the training process.

Together, these advances form a new toolbox for designing models that *know how well they've been trained* and *how much they can be trusted*.

4 - Scaling MINLPs via Branch-and-Bound and First Order Methods with Applications to Structured Statistical Learning

Rahul Mazumder, MIT, Cambridge, MA, United States

Mixed integer programming (MIP) especially, mixed integer nonlinear programming (MINLP), is a promising tool for modeling and solving key statistics/machine learning (ML) problems that admit a combinatorial description. Several statistical estimators that can be expressed as solutions to MIP problems are known to enjoy optimal statistical properties. However, they lead to computational challenges. I will discuss an approach based on an integrated branch-and-bound and first-order method framework to address a family of problems in statistics/machine learning. In a variety of learning instances, under mild statistical modeling assumptions, the underlying MIP can be simplified into another MIP which is empirically "simpler" to solve (example, MISDP can be simplified into a MIQP). The algorithms we discuss are all open-sourced.

MB06

Building A Level 3 A306

Human–AI Collaboration for Smarter Business

Invited Session

eBusiness

Chair: Tian Lu, Arizona State University, Tempe, AZ, United States

1 - Responsible Engagement: Organizational Strategy to Counteract Agency Reversal

Thomas Ware, Arizona State University, Tempe, AZ, United States

This study examines the challenge of agency reversal in AI-assisted decision-making, where human decision-makers become passive executors of algorithmic recommendations rather than active evaluators. Through a field experiment in the microlending sector, we introduce the concept of responsible engagement (RE) as a cognitive mechanism designed to counteract agency reversal without sacrificing the efficiency benefits of AI. Drawing on Work System Theory, we implement a behavioral nudge that positions decision-makers as the primary authority in the AI-assisted process. Our findings demonstrate that evaluators exposed to RE exhibit deeper information processing, more discerning integration of AI recommendations, and stronger perceptions of personal responsibility. Notably, these improvements in human agency do not come at the cost of decision efficacy or increased cognitive burden. The study contributes to our understanding of how socio-technical design can mitigate agency reversal while preserving the instrumental value of AI in operational decision-making, offering practical insights for responsible AI governance in financial and other organizational contexts.

2 - Expertise and Inevitable Errors of Algorithms

Yu Gao, Peking University, Beijing, China, People's Republic of

Growing evidence shows that experts tend to be reluctant to use artificial intelligence (AI)–based decision-support systems. This algorithm aversion poses a challenge to widespread adoption of AI decision support, given the importance of these experts' role in the diffusion process. We introduce a framework emphasizing AI's and decision makers' ability in making stable predictions. In a series of experiments, we explore how individuals, classified as "experts" and "laypeople" based on an exogenous manipulation of their prediction stability, respond to the inevitable errors made by AI. We find that experts significantly decrease reliance on AI after encountering single-instance errors, despite knowledge of its overall accuracy. Conversely, they do not increase reliance even when AI outperforms their predictions. This asymmetric response of experts persists across different tasks, even those beyond their expertise. This study advances our understanding of the emergence of algorithm aversion in decision makers with high expertise, providing valuable insights for improving AI decision-support system acceptance.

3 - Human OR Robot: Evidence from Last-Mile Delivery Service

Baorui Li, Naveen Jindal School of Management, University of Texas at Dallas, Richardson, TX, United States, Xincheng Ma, Brian Han, Daizhong Tang, Lei Fu

The increasing integration of service robots into offline environments has ushered in a transitional phase where human and robot service providers co-exist. Although prior research has explored consumer acceptance of service robots, little is known about how contextual features and underlying mechanisms shape consumer's choice in real-world settings. To address this gap, we conduct an empirical study on consumer's choice of human versus robot service, analyzing 236,776 package-level choice records from Alibaba's last-mile delivery stations. We propose that privacy sensitivity, product value, and environmental complexity elicit distinct psychological mechanisms. Our findings reveal that consumers are significantly more likely to choose robot delivery for privacy-sensitive packages (11.11%, $p < 0.001$) and high-value products (0.86% per 1% increase in value, $p < 0.001$), but prefer human couriers under adverse weather conditions (1.68%, $p < 0.001$). These findings underscore that delivery choices are shaped not only by functional considerations but also by psychological concerns, highlighting the need for context-aware service design that aligns strategies with consumer perceptions of privacy, risk, and trust.

4 - Market Intelligence and Complementors' Product Strategies: a Large-Scale Empirical Study of An E-Commerce Platform

Bowen Lou, University of Southern California, Los Angeles, CA, United States, Zhe Yuan, Tianshu Sun, Lu Fang

How does enhanced market intelligence, facilitated by platform-provided tools, shape complementors' product strategies? Collaborating with a leading e-commerce platform, we analyze over 40,000 sellers and a policy change that expanded access to a market intelligence tool. Our analysis shows that greater information access leads to more new product launches, shorter product life cycles, and higher prices for new offerings. Sellers primarily introduce products with novel attributes outside their existing portfolio, often integrating features observed from peer sellers through the tool. This expands their search space for innovation while driving convergence in customer bases. Overall, we highlight how enhanced engagement with market intelligence influences complementors' strategies through imitation, differentiation, and pricing adjustments—offering important implications for platform governance and policy.

5 - Crowding the Machine: How Does Widespread Large Language Model Adoption Reshape Profile Screening

Tian Lu, Arizona State University, Tempe, AZ, United States

The rapid rise of large language models (LLMs) has democratized professional-quality writing, enabling widespread use for tasks like profile polishing. This creates a strategic dilemma for decision platforms: should they encourage LLM use to enhance applicant communication or restrict it to preserve the effectiveness of human evaluation? The surge in LLM tutorials and services has intensified this tension. While LLMs improve surface-level writing, their adoption in high-stakes contexts—such as recruitment, admissions, and financial services—raises concerns that polished but homogenized profiles could obscure evaluators' ability to identify high-quality candidates. This study investigates crowd-level LLM use in microloan credit risk evaluation. Using a mixed-methods approach, we examine how borrowers' use of LLMs to enhance loan statements affects human evaluators' screening behavior and decision outcomes. Our findings show a non-monotonic relationship: as LLM use increases, evaluators face greater cognitive burden from diluted signals, shifting effort allocation and reducing decision quality. Building on these insights, we use analytical modeling to assess strategies for microloan platforms—whether to provide LLM toolkits to applicants or deploy LLM detectors during screening—and explore broader implications of widespread LLM use. This research contributes by revealing unintended consequences of LLM adoption in evaluative settings and identifying conditions that amplify these effects. It offers practical guidance for managing AI-human interaction in sensitive decision-making environments.

MB07

Building A Level 3 A307

Intelligent Fulfillment and Transportation

Job Market Showcase

Job Market Showcase

Chair: Kerim U. Kizil, Texas A&M University, College Station, TX, 77843, United States

1 - Towards Nature-Inspired Design and Control of Autonomous Multi-Agent Systems

Anahita Jain, The University of Texas at Austin, Austin, TX, United States, Husni Idris, John-Paul Clarke

Autonomous vehicle operations are expected to grow in the skies over the next decade, beginning with applications such as firefighting and cargo delivery, and expanding to passenger transport. These vehicles are expected to interact and integrate with other traffic within the airspace. For scalability, this airspace of the future will be a collective system of autonomous vehicles, where each vehicle makes increasingly independent decisions.

The successful implementation of these decentralized, autonomous, collective multi-vehicle systems can benefit from understanding existing multi-agent systems, such as colonies of ants or schools of fish. This helps us identify the *characteristics* or *behaviors* that make these systems “good”, and subsequently identify the metrics that *quantify* these characteristics. Further, these metrics can be leveraged for the design of agent behaviors and coordination mechanisms to *achieve* these desired traits in the collective autonomous multi-agent system.

We hypothesize that at low densities, “self-separation” by aircraft maintaining minimum separation distances is sufficient to ensure safety. As density increases, “self-organization” becomes necessary for maintaining efficiency. This is achieved by enabling aircraft to independently and dynamically create airspace order when beneficial, by adjusting the degree to which they follow other airspace traffic. At even higher densities, “self-limitation”- using metrics such as trajectory flexibility- becomes necessary to ensure aircraft have enough room to maneuver to mitigate the effects of uncertainties and perturbations.

In this talk, we will explore each of these self-behaviors and their interrelationships, using the nature-inspired models and capabilities we developed to enable them in aviation and other multi-agent systems.

2 - Physics-Informed Machine Learning for Enhanced Traffic Control: from Megavandertest to Neural Finite Volume Methods

Zhe Fu, UC Berkeley, Berkeley, CA, United States

Strategically controlling a small number of “leader” vehicles can significantly enhance traffic flow, improving efficiency, fuel economy, and driver comfort for the entire traffic system with minimal intervention. This was validated in the MegaVanderTest, a large-scale field experiment involving 100 autonomous vehicles (AVs) operating in mixed autonomy traffic. We designed both model-based and machine learning-driven control strategies for the AVs to improve traffic congestion and reduce the system-wide fuel consumption by over 10%.

However, the effectiveness of these control strategies heavily relies on accurate traffic state predictions and reliable communication. To address these challenges, we developed the Neural Finite Volume Method (NFVM), a physics-informed machine learning framework that integrates neural networks with physics-based models to solve hyperbolic partial differential equations (PDEs). NFVM achieves high predictive accuracy while maintaining key physical principles such as conservation laws and entropy conditions, making it well-suited to capturing complex traffic phenomena like shocks and stop-and-go waves. NFVM not only outperforms traditional finite volume solvers in accuracy but does so with comparable computational cost. It also offers structural flexibility, enabling multi-timestep inputs and the ability to generalize from PDE-synthetic data to experimental field data.

By embedding NFVM as the traffic state estimation layer within the AV control stack, we further enhance the performance and robustness of our control strategies.

3 - UAV Search and Routing Planning in a Disaster Area

Nastaran Olad zad-Abbasabady, University at Buffalo-SUNY, Amherst, NY, United States, Esther Jose, Rajan Batta, Miguel Lejeune

The goal of this work is to identify the best Search and Routing Planning (SRP) for an Unmanned Aerial Vehicle (UAV) to maximize the total number of casualties detected in a disaster-affected area within a limited mission duration. First, we introduce a Mixed-Integer Nonlinear Programming (MINLP) formulation with a nonlinear objective function. To linearize the objective function, additional assumptions and constraints are imposed, resulting in a Mixed-Integer Linear Programming (MILP) approximation of the original problem. We then describe an exact solution method and a hybrid heuristic that yield higher-quality solutions. In addition, a clustering-based heuristic is proposed to handle larger problem instances. The framework is extended to account for uncertainty in casualty estimates and coordination among multiple UAVs. The effectiveness and computational performance of the proposed methods are validated through a comprehensive set of experiments. Finally, a case study inspired by the 2023 Turkey-Syria earthquakes demonstrates the real-world applicability of our approach to time-critical post-disaster search operations.

4 - Robotic Mobile Fulfillment Systems: Strategies for Pod Selection and Scheduling

Kerim U. Kizil, Texas A&M University, College Station, TX, United States, Chelliah Sriskandarajah, Jon Stauffer

Robotic mobile fulfillment (RMF) systems automate storage and transportation within fulfillment centers while still relying on human pickers. We focus on two performance metrics in these systems, aiming to minimize overall completion time (OCT) as the primary objective and the number of required robots (NRR) as a secondary objective. We investigate two interrelated operational problems influencing these metrics: (i) pod selection and (ii) pod scheduling. We first explore the pod scheduling problem independently, providing theoretical results. This stand-alone problem is NP-hard with at least two pickers when minimizing OCT and remains NP-hard with even one picker when minimizing NRR. The NRR objective introduces a novel optimization problem structure, contributing to scheduling theory. We demonstrate that a simple but effective scheduling rule is asymptotically optimal for minimizing OCT with multiple pickers. For NRR minimization with a single picker, we derive theoretical performance bounds for two sequencing rules. When incorporating the pod selection problem, we focus on two approaches: (i) sequential and (ii) integrated pod selection and scheduling. The sequential approach handles pod selection first, then pod scheduling, while the integrated approach addresses both problems simultaneously within a single formulation to minimize OCT. Computational experiments using e-commerce order data reveal that the sequential approach achieves

significantly faster computation times, with a mean OCT similar to that of the integrated approach. We also find that our sequencing rules can reduce mean NRR by 13% to 28% without affecting OCT, compared to the case where no such rules are used.

MB08

Building A Level 3 A310

Platform Strategy and Governance

Invited Session

Information Systems

Chair: Martin Quinn, Erasmus University Rotterdam, ROTTERDAM, Netherlands

1 - Regulation of Algorithmic Collusion from Data

Chenhao Zhang, Northwestern University, Evanston, IL, United States

Algorithms are increasingly being used to price goods and services in competitive markets. Many recent papers have shown that in various settings, some configurations of certain pricing algorithms can find and maintain supra-competitive prices from repeated market interactions. We give a definition of plausible algorithmic non-collusion for pricing algorithms. On one hand, the definition allows sellers to utilize useful side information that may be correlated with supply and demand. On the other hand, it allows a regulator to empirically audit algorithms by applying a statistical test to the data they collect. We provide an analysis of the statistical complexity of such an audit, i.e., how much data is sufficient for the test of non-collusion to be accurate. We further discuss the unrelaxability of the definition even when there is no side information, as well as the role of the regulator's knowledge of the seller's cost in the audit.

2 - Decision-Making Under Uncertainty: Improving Multi-Armed Bandits and Pessimism Traps

Kavya Ravichandran, Toyota Technological Institute at Chicago, Chicago, IL, United States

We study decision-making under uncertainty in two settings: first, we give nearly-tight upper and lower bounds for the improving multi-armed bandits problem. An instance of this problem has k arms, each of whose reward function is a concave and increasing function of the number of times that arm has been pulled so far. We show that for any randomized online algorithm, there exists an instance on which it must suffer at least an $\Omega(\sqrt{k})$ approximation factor relative to the optimal reward. We then provide a randomized online algorithm that guarantees an $O(\sqrt{k})$ approximation factor, if it is told the maximum reward achievable by the optimal arm in advance and $O(\sqrt{k} \log k)$ approximation if it must estimate that quantity. In the second work, we model pessimism traps -- a phenomenon studied by philosophers in which agents are influenced by their communities to not engage in ambitious ends. Adapting an opinion dynamics model to capture the key facets of pessimism traps, we show how easily these traps develop. In order to sustainably shift a community out of pessimism, we develop a targeted subsidy scheme. We then extend such an intervention to a setting where there are many different communities for whom different actions are considered optimistic, showing that even with limited information about the communities, a centralized agency can help shift all of them toward optimism.

3 - Boosting in-Context Learning in LLMs Through the Lens of Classical Supervised Learning

Korel Gundem, George Washington University, Washington, DC, United States, Juncheng Dong, Dennis Zhang, Vahid Tarokh, Zhengling Qi

In-Context Learning (ICL) allows Large Language Models (LLM) to adapt to new tasks with just a few examples, but their predictions often suffer from systematic biases, leading to unstable performances in classification. While calibration techniques are proposed to mitigate these biases, we show that, in the logit space, many of these methods are equivalent to merely shifting the LLM's decision boundary without having the ability to alter its orientation. This proves inadequate when biases cause the LLM to be severely misdirected. To address these limitations and provide a unifying framework, we propose Supervised Calibration (SC), a loss-minimization based framework, which learns an optimal, per-class affine transformation of LLM's predictive probabilities in the logit space without requiring external data beyond the context. By using a more expressive functional class, SC not only subsumes many existing calibration methods in ICL as special cases but also enables the ability of altering and even completely reversing the orientation of the LLM's decision boundary. Furthermore, SC's loss-based nature facilitates the seamless integration of two purpose-built regularization techniques—context-invariance and directional trust-region regularizers. The former is designed to tackle the instability issue in ICL, while the latter is to control the degree of calibration. Finally, SC delivers state-of-the-art performance over calibration baselines in the 4-shot, 8-shot, and 16-shot settings across all nine datasets for Mistral-7B-Instruct-v0.3, Llama-2-7B-chat, and Qwen2-7B-Instruct.

4 - Generative AI and the Division of Labor in Knowledge Work

Maximilian Schaefer, Institut Mines-Télécom Business School, Evry, France, Vincent Lefrere

This study investigates the impact of Large Language Models (LLMs), such as GitHub Copilot, on software development practices. We construct a longitudinal panel dataset of GitHub developers from 2011 to 2025 using event-level data and API-enriched metadata. The dataset captures detailed information on code contributions, collaboration patterns, and developer profiles. Our research focuses on two main questions: (1) how LLM adoption influences productivity, code reuse, and team structure, and (2) how the presence of AI-generated code can be identified using watermarking and machine learning-based detection methods. By linking developer behavior to the diffusion of generative AI tools, we contribute to a broader understanding of how automation transforms knowledge work and collaborative production at scale.

MB09

Building A Level 3 A311

Information Disclosure and Transparency

Job Market Showcase

Job Market Showcase

Chair: Ruru Hoong, Harvard Business School, Cambridge, MA, 02138, United States

1 - Feedback Design and Equilibrium Knowledge-Sharing

Xinbo Wang, National University of Singapore, Singapore, Singapore, Nan Chen

Feedback systems are widely used by digital platforms. We study how feedback design shapes equilibrium content provision in the context of Stack Overflow. Using a regression discontinuity design based on answer timing, we show that competition shapes feedback and influences contribution. We develop and estimate an empirical model that links feedback design to equilibrium outcomes among heterogeneous contributors. The model allows the platform to continuously adjust *leniency* (the tendency to give positive feedback) and *harshness* (the tendency to give negative feedback). We decompose feedback's impact into a direct effect on individual contributors and an indirect competitive effect. We find that although a contributor favors more leniency and less harshness when considered alone, she may prefer less leniency and higher harshness in equilibrium. This reversal arises from competition and depends on user quality, with higher-quality contributors being more likely to alter their preferences under equilibrium incentives. We further show how the platform's targeted quality levels, its ability to distinguish contributor quality, and the distribution of contributor skill shape competition and inform optimal feedback design. Finally, we consider the case of endogenous feedback preferences, where contributors' valuations of feedback depend on the equilibrium quality induced by the mechanism.

2 - How Platform Transparency Shapes Provider Choices: Evidence from a Natural Experiment on Lyft

Rubing Li, New York University, New York, NY, United States

In February 2024, Lyft introduced a policy guaranteeing drivers at least 70% of rider payments while also increasing per-ride earnings transparency. The rollout of this policy was staggered, first introduced in "major markets" that were more urban, and provided a natural experiment to assess how platform transparency and earnings guarantees affect ridesharing availability, driver engagement and rider satisfaction. Using trip-level data from over 47 million rides in a major urban market and its neighboring suburban markets across six months, we applied dynamic staggered difference-in-differences models to measure the causal effects of these platform design changes on supply- and demand-side outcomes, and ensuing ridesharing performance. We show that the design change led to substantial changes in driver engagement, with separate effects from the guarantee and the transparency. Drivers increased their working hours and utilization, leading to more completed trips and higher per-hour

and per-trip earnings. These effects were strongest for drivers with lower pre-policy earnings and greater income uncertainty. We unpack the economic mechanism by which increased supply also had a positive spillover on demand, boosting rider app engagement and booking conversions and lowering wait times in high demand areas. We also provide some evidence that points to platform transparency potentially leading to unintended strategic driver behavior.

3 - Sharing the Spotlight: the Impact of a Credit System on Creator Collaboration Dynamics in An Online Platform

Yue Zhang, The Hong Kong University of Science and Technology, Hong Kong, China, People's Republic of

Collaboration yields high-quality outcomes in business, research, and online content creation. However, collaboration can be hindered when individual contributions go unrecognized. In this paper, we study how a formal credit system in an online video platform affects creators' collaboration dynamics and examine how these effects vary across creators of different popularity levels. We find that implementing the credit system significantly increases the likelihood that a video is produced collaboratively, as well as the engagement metrics of videos that adopt the system. We further demonstrate that the system's impact increases with a creator's popularity but has little effect on less popular creators. This raises concerns about unintended consequences, as the credit system may exacerbate attention disparities within the creator community. Additionally, we analyze how the credit system influences collaboration patterns between creators and their collaborators. We find that creators are more likely to collaborate with creators of similar popularity, direct competitors, and previous collaborators. Our study contributes theoretically by elucidating the interplay among credit attribution, collaboration, and competition on online platforms, and offers practical insights for platform owners to balance the benefits and costs of implementing a formal credit system.

4 - Calibrated Coarsening: Designing Information for AI-Assisted Decisions

Ruru Hoong, Harvard Business School / Harvard Economics, Cambridge, MA, United States

Artificial intelligence (AI) signals are increasingly deployed as human decision-making aids across many critical applications, but human cognitive biases can prevent them from improving outcomes. We propose *calibrated coarsening*—partitioning the signal space into fewer cells at optimised thresholds—as a way to improve decision-making outcomes while (i) keeping humans in the loop, (ii) modifying signals without deception, and (iii) adapting flexibly to various cognitive biases and decision-making contexts. Within an optimal information disclosure framework, we derive the approximately-optimal universal coarsened policy for settings where the designer does not observe the decision-maker's information. We then empirically demonstrate in a randomised experiment involving loan specialists that coarsening AI signals at the theory-derived threshold significantly improves decision-making outcomes, over both the human-only (based solely on the loan application) and continuous AI (assisted with uncoarsened AI risk-score) benchmarks. We uncover substantial decision heterogeneity amongst loan officers, and use a Bayesian hierarchical model to personalise coarsening policies, which can further improve outcomes as past data become available.

MB10

Building A Level 3 A312

Sustainable System Design

Job Market Showcase

Job Market Showcase

Chair: Parisa Shahsavand, The University of Texas at Dallas, Richardson, TX, United States

1 - Ranking Crop Cultivars Using Stochastic Dominance

Shayan Tohidi, Iowa State University, Ames, IA, United States, Sigurdur Olafsson

Sustainability of food production relies on crop cultivars that perform well under a range of environmental conditions. This requires comparing crop cultivars as random prospects, where the uncertainty is due to the environment. Stochastic Dominance (SD) is a classical way of comparing and ranking random prospects, which accounts for the outcome and risk. The goal is to find those with Maximum Expected Utility (MEU), and the SD rules guarantee the results, e.g., the first and second order stochastic dominance do it for all and risk-averse decision makers, respectively. While these methods are important, they may not apply in practice due to existing extreme utilities that enlarge the efficient set. Almost Stochastic Dominance (ASD) has been proposed to eliminate extreme utility functions and be acceptable to most decision makers. Although it shrinks the size of the efficient set, its magnitude depends on a tunable parameter ϵ . Moreover, the dominant prospect always should have a higher expected value, even if that margin is small. This makes ASD less applicable in certain real-world problems, such as crop cultivar selection, where it may be sensible to prefer a more stable cultivar to a cultivar with slightly higher expected yield. We propose a new heuristic method that does not require tunable parameters and has no constraints on the expected value. The relationship with other SD and ASD methods and the special cases is described, and the method is applied to a crop selection case study.

2 - Optimizing Distributed Manufacturing of Biomass-Based Chemicals with Electrochemistry and Renewable Energy

Motahareh Kashanian, Iowa State University, Ames, IA, United States

The push for sustainable chemical manufacturing has sparked interest in distributed production systems powered by renewable energy. This talk presents a multi-study framework for optimizing the distributed manufacturing of biomass-based chemicals, focusing on electrochemical processes as a greener alternative to traditional methods. We explore the integration of local renewable energy sources, solar and wind, into electrochemical conversion systems, aiming to reduce greenhouse gas emissions and improve the economic viability of decentralized production.

Our approach combines supply chain design, techno-economic analysis, and uncertainty modeling to evaluate the feasibility of these systems at scale. Using mixed-integer linear programming and stochastic programming, we model the trade-offs between centralized and distributed production under various policy, technical, and environmental scenarios. We also investigate how process scale, renewable availability, and the social cost of carbon influence facility location, size, and overall system performance.

Results highlight the role of modular, renewable-powered systems in lowering emissions while maintaining cost-effectiveness, especially under dynamic energy and policy conditions. The framework is demonstrated through case studies for the production of trans-3-hexenedioic acid (t3HDA), offering insights into real-world applicability and pathways for broader deployment.

3 - Systems Modeling and Analysis for a Sustainable Material Future

Sidi Deng, University of Michigan, Ann Arbor, MI, United States, Daniel Cooper, Yuehwen Yih, John Sutherland

Materials are the building blocks of industrial society, flowing through supply chains that represent complex, interconnected networks. Given this, material sustainability is inherently a systemic challenge that calls for cross-cutting, macro-oriented strategies grounded in systems thinking. This talk presents an interdisciplinary research roadmap for fostering a sustainable, material-centered industrial landscape through the lens of system modeling and analysis. Central to this job talk is how model-based virtual representations of industrial ecosystems, combined with data-driven approaches, inform decision-making to advance key sustainability dimensions, including material circularity, economic feasibility, environmental performance, and supply chain resilience. The talk opens by highlighting the presenter's Ph.D. work on advancing the circular economy of critical materials essential to the clean energy transition, incorporating and synergizing a diverse set of computational tools, including simulation, machine learning, network analysis, and system dynamics. The middle section highlights the presenter's current research on decarbonizing the aluminum industry through the modeling and simulation of supply chains and technology pathways. At its core is an agent-based computational engine that drives Monte Carlo simulation, dynamic material flow analysis, life cycle assessment, and risk-informed scenario analysis using experimental design. The job talk concludes with the presenter's future vision for navigating complex industrial systems, highlighting opportunities to integrate geospatial analysis, physics-informed machine learning, and digital twins. Through his research, the presenter seeks to drive the transition toward a smart and resilient industrial landscape rooted in sustainable material systems.

4 - It's Still Good! Expiration Dates and Food Waste

Parisa Shahsavand, The University of Texas at Dallas, Richardson, TX, United States, Dorothee Honhon, Iva Rashkova

This paper examines how expiration dates influence food waste and consumer behavior. In the United States, expiration dates (except for infant formula) are not federally regulated and are typically intended to reflect product quality, not safety. However, many consumers misinterpret them, leading to premature disposal of still-edible food and contributing substantially to household food waste. We model a food producer that determines the shelf life and production cycle of a perishable product and analyze consumer behavior along two dimensions: whether they consider expiration dates at the time of purchase and whether they consider them when consuming products at home. Our results show that when consumers ignore expiration dates during purchase, producers tend to set both the shelf life and production frequency equal to the biological life of the product, resulting in higher absolute waste at the consumer level. Most notably, we find that food waste is minimized when consumers pay attention to expiration dates while shopping but are willing to consume products past the expiration date at home, as long as the product remains palatable. This consumer behavior allows for informed purchasing and reduced unnecessary disposal. Our findings suggest that expiration dates, when properly understood and used, can support rather than hinder waste reduction efforts. Consumer education that frames expiration dates as indicators of quality rather than safety can promote more sustainable consumption behavior and better alignment between producer decisions and societal waste reduction goals.

MB11

Building A Level 3 A313

Industrial Reliability and Maintenance

Job Market Showcase

Job Market Showcase

Chair: Anyi Li, Auburn University, Auburn, AL, 36830, United States

1 - Physics- and Shape-Constrained Data Science Methods with Applications in Materials and Industrial Informatics

Yating Fang, Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, United States, Qianqian Zhao, Ryan B. Sills, Ahmed Aziz Ezzat

This talk highlights my dissertation research on scientific machine learning for materials and industrial systems. I will present a shape-constrained method that introduces, for the first time, monotonicity constraints to the mixture of experts framework (a prevalent ensemble learning method), with applications in materials failure and asset management. This is followed by an overview of my related research and educational accomplishments and future agenda.

2 - Advancing Additive Manufacturing with Physics Informed Deep Learning: Geopipn for Fast Thermal Prediction in Unseen Part Geometries

Amirul Islam Saimon, Virginia Tech, Blacksburg, VA, United States, Raghav Gnanasambandam, Zhenyu Kong

This talk will briefly review current progress and key challenges in applying deep learning to additive manufacturing. It will then introduce Geometry-Aware Physics Informed PointNet (GeoPIPn), a novel deep learning-based fast prediction model for solving Partial Differential Equations (PDE) across unseen geometries. Thermal distributions in additive manufacturing (AM) parts, governed by diffusion PDEs, are key to ensuring material quality, microstructure uniformity, and dimensional accuracy. Existing thermal modeling relies on Finite Element Method (FEM), which is computationally expensive for fast-paced, iterative geometric design. Furthermore, these methods are tailored to a specific part geometry, requiring rerunning for each new geometry, making them inefficient for designs that change frequently. Once trained, our GeoPIPn can predict the thermal distribution for a new part geometry very quickly without rerunning the FEM simulations. In the case studies, we demonstrate the superior performance of our GeoPIPn over state-of-the-art benchmark methods. Remarkably, GeoPIPn is the first PINN-based model focusing on thermal modeling in AM for new part geometries with improved performance. Consequently, it offers a powerful solution for fast and accurate AM design.

3 - Process-Driven Machine Learning for Intelligent and Collaborative Fatigue Modeling in Additive Manufacturing

Anyi Li, Auburn University, Auburn, AL, United States, Jia Liu

Understanding the fatigue behavior and accurately predicting the fatigue life of laser power bed fusion (L-PBF) parts remains a pressing challenge due to complex failure mechanisms, time-consuming tests, and limited fatigue data. My research tackles these challenges through two key directions: (1) developing a physics-informed, nondestructive fatigue life prediction framework, and (2) designing a generalized, privacy-preserving federated learning model that supports cross-organizational collaboration in heterogeneous and data-scarce environments.

The first contribution introduces a Multimodal Transfer Learning (MMTL) framework to understand process-defect-fatigue relationships in L-PBF parts via a hierarchical graph-based representation. MMTL leverages a pre-trained model with abundant process and defect data in the source task to predict fatigue life nondestructively with limited fatigue test data in the target task. This work demonstrates the feasibility of nondestructive fatigue life prediction even with limited labeled data, offering a scalable and interpretable solution to enhance AM reliability.

The second presents a personalized federated learning framework that allows multiple manufacturers to collaboratively train predictive models without sharing their raw data. To address data heterogeneity, especially clients with differing feature sets-we incorporate domain adaptation and statistical alignment techniques to share knowledge from feature-rich clients to those with missing features. This framework enables small and medium-sized manufacturers to benefit from collective intelligence while maintaining data privacy and operational independence.

MB12

Building A Level 3 A314

Designing and Evaluating Environmental Policy Instruments

Job Market Showcase

Job Market Showcase

Chair: Shuo Yu, Berkeley, 94720, United States

1 - Optimal Subsidies for Carbon Capture: a Stackelberg Game Analysis

Connor Colombe, University of Texas at Austin, Austin, TX, United States

Model-based scenarios for achieving decarbonization tend to envision significant deployment of carbon capture (CC) technologies. In the United States, the federal government subsidizes CC through its 45Q tax credits. In this paper, we develop and analyze a Stackelberg game model featuring a social-welfare-maximizing government and a profit-maximizing firm. The government sets the level of a CC subsidy and the firm responds by jointly deciding whether or not to invest in CC technology and on its production level. In the case where the government has perfect information about the firm's parameters, we analytically derive the optimal CC subsidy and establish the conditions under which inducing the firm to invest in CC will actually increase carbon emissions. We then consider an extension where the firm's CC investment cost is uncertain from the government's perspective. In this case, we show that uncertainty can situationally raise or lower the optimal subsidy. Lastly, we apply our model to a numerical case study of a coal power plant and estimate optimal subsidy values. We find that the current subsidy level of \$85 Mt/CO₂ is likely a sufficient subsidy level to incentivize CC adoption and that the value may be approximately social welfare maximizing. In addition, we find that uncertainty may cause the Government to overspend \$1.4 billion over a ten year period when compared to the subsidy offered under perfect information. Furthermore, we find that the subsidy value our model predicts led to an 83% reduction in CO₂ emissions when compared to the no-subsidy equilibrium value.

2 - Optimal Policy Portfolios to Promote Carbon Capture

Abdullah Albeladi, The University of Texas at Austin, Austin, TX, United States, Benjamin Leibowicz

Carbon capture (CC) technologies are projected to play important roles in deep decarbonization pathways, yet their deployment to date has been limited. Current policy incentives for CC primarily take the form of flow subsidies per ton of carbon dioxide (CO₂) captured, as in the 45Q tax credits in the U.S. In this paper, we investigate whether a policy portfolio consisting of multiple instruments -- a flow subsidy, an upfront investment subsidy, and an emissions penalty -- could spur CC adoption, reduce net CO₂ emissions, and contain government costs more effectively than the prevailing flow-subsidy-only approach. We formulate a Stackelberg game model in which the government sets the levels of these policy instruments to maximize social welfare, anticipating the reactions of heterogeneous firms who choose whether to adopt CC, their production levels, and CO₂ capture rates to maximize profits. After deriving analytical solutions to the firms' problems and for the share of firms who adopt CC, we apply the model to a numerical case study of natural gas-fired power plants. We find that the optimal policy design employs all three instruments and yields 52% higher social welfare than a flow subsidy alone. Our results clarify the primary role that each instrument should fulfill in the optimal portfolio: the upfront subsidy induces firms to install CC, the flow subsidy covers baseline CC operating costs, and the emissions penalty incentivizes high CO₂ capture rates.

3 - Carbon Capture and the Power Sector: Health and Environmental Outcomes of US Policies

Paola Furlanetto, UMass Amherst, Amherst, MA, United States, Erin Baker, Nexus Attiogbe, Michael Ash

Recent U.S. policies—such as the Inflation Reduction Act's 45Q tax credit and the EPA's CO₂ Performance Standards for fossil fuel power plants—seek to accelerate carbon capture (CC) deployment in the power sector. This study explores the interactions between these policies, the EPA's Good Neighbor rule, and a hypothetical carbon tax across diverse technical and geographic scenarios. Using a power flow model of a stylized three-node grid, we simulate investment decisions in CC retrofits and new generation capacity, while tracking emissions of both CO₂ and harmful co-pollutants (NO_x, SO_x, PM_{2.5}, NH₃ and VOCs).

Our analysis reveals that while CC incentives can reduce greenhouse gas emissions, they may also unintentionally promote continued fossil fuel use and delay renewable energy deployment. In some scenarios, co-pollutant emissions and associated health damages increase, though these can be mitigated through complementary policies that directly regulate co-pollutants. These findings highlight the need for integrated policy design that considers not only carbon outcomes but also public health. As the U.S. navigates a rapidly evolving climate policy landscape, our results underscore the importance of aligning decarbonization strategies with broader societal goals.

4 - Evaluating Payments for Ecosystem Services: Cover Cropping and Water Quality in the Midwest

Shuo Yu, University of California, Berkeley, Berkeley, CA, United States

Agricultural runoff significantly contributes to nutrient water pollution, leading to harmful environmental and economic consequences. Cover cropping (CC), the practice of planting non-cash crops during off-seasons, has emerged as a promising conservation strategy to mitigate these impacts. Payment for ecosystem services programs, such as the USDA's Environmental Quality Incentives Program (EQIP), have increasingly supported CC adoption through incentives. This study evaluates the social value of EQIP in enhancing water quality by examining its impact on CC adoption and the subsequent reduction in water pollution in the Midwest. Leveraging a unique 17-year, satellite-derived dataset on plot-level CC adoption and granular harmonized water quality metrics, we find that a one percentage point increase in upstream CC adoption reduces total nitrogen in surface water by 0.9%. An event-study analysis of EQIP's Mississippi River Basin Initiative shows an initial 8.7 percentage point increase in CC adoption, with the impact diminishing over time. Overall, EQIP's CC practices recover 48% of their implementation costs through reductions in nitrogen pollution under conservative assumptions, suggesting partial but significant economic returns via water quality improvements.

MB13

Building A Level 3 A315

Academic Job Search Panel

Panel Session

Career Fair

Co-Chair: Priyank Arora, University of South Carolina, Columbia, SC, 29208, United States

1 - Moderator Panelist

Priyank Arora, University of South Carolina, Columbia, SC, United States

The purpose of this session is to bring visibility to the students, and postdocs looking for academic positions. Panelists from both business and engineering schools will share their experiences. This panel discusses the academic interview process and do's and don'ts associated with the job search.

2 - Panelist

Lavanya Marla, University of Illinois at Urbana-Champaign, Urbana, IL, United States

3 - Panelist

Alex Mills, Baruch College, New York, NY, United States

4 - Panelist

Juba Ziani, Georgia Tech, Atlanta, GA, United States

5 - Panelist

Simin Li, Tulane University, New Orleans, LA, United States

MB15

Building A Level 4 A401

Intelligent and Safe Autonomy for Aerial Systems

Invited Session

Air Transportation Section

Chair: Yuwei Wu, University of Pennsylvania, Philadelphia, PA, United States

1 - Distributed Optimization for Safe Multi-Agent Aerial Autonomy

Matthew Hale, Georgia Institute of Technology, Atlanta, GA, United States

The future of aerial transportation is likely to involve large numbers of vehicles that must safely navigate, perhaps in close proximity. Close proximity means that collision avoidance is of paramount importance, yet the presence of large numbers of vehicles means that centralized planning may be computationally inefficient or outright infeasible. Therefore, in this talk I will present a decentralized framework for safe multi-agent aerial autonomy. First I will present recent developments on an optimization algorithm for safe, decentralized motion planning that computes optimal paths while obeying all collision avoidance constraints. This algorithm is decentralized by design, and I will show that it is robust to asynchrony in the sense that agents can asynchronously compute and communicate, and they will still find a safe, optimal motion plan. Then I will demonstrate the results of such plans experimentally on aerial vehicles. These experiments show the success of this framework in practice, and they demonstrate that even computationally constrained, lightweight, and low-power aerial agents can reliably compute safe, optimal motion plans.

2 - Llm-Based End-to-End Flight Planning with Integrated Multimodal Review

Amin Tabrizian, George Washington University, Washington, DC, United States

We propose a novel end-to-end framework for automated flight planning that leverages large language models (LLMs) in both the planning and review stages, incorporating visual inputs such as airspace charts to enhance situational understanding. Our system, operates in two phases: (1) a multimodal LLM generates flight plans

conditioned on textual mission requirements, and (2) a secondary LLM acts as a self-review agent, evaluating the generated plans for feasibility, safety, and regulatory compliance.

To enable multimodal reasoning, we leverage a VLM that integrates spatial features with domain-specific flight constraints, allowing it to generate route suggestions, waypoint justifications, and alternate plans flight planning scenarios. We validate our approach using simulated flight scenarios sourced from aviation datasets. Quantitative metrics show significant improvements in route optimality, safety constraint satisfaction, and interpretability compared to traditional planning pipelines. Moreover, qualitative results highlight the coach LLM's ability to detect subtle errors, such as misaligned airspace boundaries or infeasible altitudes.

This work demonstrates the promise of a closed-loop LLM framework for safe and efficient autonomous flight planning, and opens new directions for human-AI collaboration in safety-critical domains. We release all models, prompts, and evaluation tools to facilitate future research.

3 - Implicit Learning of Riemannian Polynomials for Quadrotor Trajectory Optimization

Yuwei Wu, University of Pennsylvania, Philadelphia, PA, United States

This talk addresses the problem of generating optimal trajectories on Lie Groups with left-invariant Riemannian metrics, a challenge that traditionally requires solving nonlinear boundary value problems through numerical methods. We develop a coordinate-free framework to learn optimal smooth trajectories using implicit neural networks with an arbitrary initial and final state, including non-homogeneous boundary conditions with any non-rest start and end state. We parameterize the trajectory via higher-order polynomials and construct a loss function that incorporates the necessary conditions for optimality and boundary conditions. The proposed framework is evaluated against numerical optimization methods, and the optimality is analyzed to validate the proposed approach. Furthermore, we demonstrate the capacity of our method to generate piecewise trajectories through multiple waypoints, which can be efficiently integrated into quadrotor trajectory planning and control frameworks.

4 - Real-Time Safe and Energy-Efficient Uav Motion Planning in Windy Urban Environments

Spencer Folk, University of Pennsylvania, Philadelphia, PA, United States

Recent advancements in hardware and software are bringing autonomous aerial vehicles closer than ever to delivering futuristic visions of flying cars and package delivery drones. However, the safe deployment of autonomous aircraft at scale in urban environments poses significant challenges, one of which being uncertainties contributed by complex wind fields. Building clusters generate dangerous high speed wind flow patterns which increase both the costs and risks for aerial operations. Existing solutions rely on either distributed sensors, extensive exploration, or accurate global representations of the environment with expensive offline calculations.

In this talk, we first introduce a novel approach to predicting wind flow fields in real time, without GPS or prior knowledge of the environment, and only using on-board sensing. We accomplish this by strategically reducing the problem by predicting the wind flow field on a local robot-centered domain and thus leveraging the predictability of the wind flow field induced by conservation of mass and momentum constraints from nearby topography. Using vast amounts of simulated winds through procedurally generated urban environments, we train deep neural networks to synthesize navigational LiDAR scans with sparse in-situ measurements of the wind, achieving real-time prediction rates and surprising generalizability. In the latter half of the talk, we demonstrate how the local wind prediction can be incorporated into a receding horizon optimal control architecture to reduce crash rates and energy consumption. The combined system can perceive surrounding winds in real time with on-board sensors and adjust the planned route on the fly.

5 - Enabling Formal Guarantees in Learning-Based Autonomy with Challenging Dynamics

Shreyas Kousik, Georgia Institute of Technology, Atlanta, GA, United States

As robots become more widespread through society, we have an immense opportunity to improve their performance and integration through massively parallelized scaling of data collection and learning, we are already seeing this with autonomous driving. At the same time, this means it is absolutely critical to ensure that such learning-enabled robots are safe, especially in aviation. To this end, this talk presents our recent efforts in learning of and with formal guarantees. The key idea is that safety and performance are not necessarily a tradeoff, as one might initially suspect. Careful design of both learning and deployment can in fact get the best of both worlds.

MB16

Building A Level 4 A402

Prescriptive and Predictive Analytics in Logistics and Supply Chain

Invited Session

The Practice Section of Informs

Chair: Nitish Umang, Johnson & Johnson, Clifton Park, NY, United States

1 - Weekly Home Healthcare Routing and Scheduling with Overlapping Patient Clusters

Ai Zhao, Johnson & Johnson, Plano, TX, United States

We present a two-step approach for efficiently solving a weekly home healthcare scheduling and routing problem. Two new MIP models are proposed, where the first is used for making patient-therapist assignments over the week, and the second for deriving daily routes. In both MIPs, the objective function contains a hierarchically weighted set of goals. The major components of the full problem are continuity of care, downgrading, workload balance, time windows, overtime, and mileage costs. A new preprocessing procedure is developed to limit the service area of each therapist to a single group of overlapping patients. Once the groups are formed, weekly schedules are constructed with the MIPs. The overall objective is to minimize the number of unscheduled visits and total travel and service costs subject to the operational and physical constraints mentioned above. Computational experiments are conducted with real data sets provided by a national home health agency. The results show that optimal solutions can be obtained quickly for large-size instances, and that they compare favorably with results obtained with a proposed integrated model as well as the actual schedules.

2 - AI-Based Smart Inspection and Anomaly Detection in Supply Chain

Reyhaneh Nosrati, Johnson & Johnson, Raritan, NJ, United States, Sandhya Kamineni, Anu Ahuja, Rami Musa, Saman Sarraf

This work presents two machine learning-based solutions designed to automate inspection processes for Johnson & Johnson products, demonstrating how Machine Learning can enhance quality assurance and operational efficiency.

The first, Automated Staple Inspection (ASI), focuses on the Ethicon EndoCutter stapler, a crucial surgical device. Currently, staple quality is assessed manually on small samples, producing binary Pass/Fail results that lack detail. The proposed ML system employs advanced computer vision techniques to perform highly granular inspections, achieving over 99% accuracy in staple form detection, height estimation, and outline measurement—surpassing KPI thresholds of 95%. This enables quantitative inspection, reduces manual reliance, and prevents revenue loss. The solution's stability and scalability, combined with cost benefits, project savings of \$230 million over ten years.

The second solution, AI-based Smart Inspection (AIS), addresses the verification of orthopedic joint reconstruction surgical trays. These trays contain numerous components arranged in designated slots, requiring precise presence and placement checks. Replacing manual inspection, the AI pipeline employs a comprehensive model capable of identifying up to 3,145 classes with over 90% accuracy. Its scalable architecture minimizes maintenance costs while ensuring high inspection fidelity, streamlining quality workflows and reducing operational costs.

Together, these solutions exemplify Johnson & Johnson's leadership in AI-driven automation, delivering high-precision, scalable, and cost-effective inspection methods. These innovations significantly improve product safety, quality, and operational efficiency in medical device manufacturing.

3 - Multi-Product 3D Container Load Planning in Outbound Logistics Operations

Nitish Umang, Johnson & Johnson, Clifton Park, NY, United States

The multi-product three-dimensional container loading problem (3D-CLP) involves packing heterogeneous cases into a three-dimensional space in a space-efficient manner. This problem has broad applications in supply chain and logistics planning across various industries where physical products must be loaded and packed for storage and/or shipment to different destination locations. The importance of 3D-CLP is evident in its applicability across a range of physical loading processes, such as the loading of multiple shipping cases from one or more product categories into a gaylord container or onto a pallet, and the subsequent loading of multiple gaylords or pallets into a transport container, such as an ocean container or a truck trailer.

Generating high-quality loading plans in three-dimensional space within reasonable computation times can be especially challenging in many commonly encountered real-world use cases. We developed a decision-aid tool called LoadMax that leverages optimization-based techniques to address this challenge of balancing solution quality and computational efficiency for outbound operational planning. The tool has been implemented and deployed to create highly efficient loading plans for gaylords and trucks used in daily operations at a major distribution logistics site at Johnson & Johnson, leading to significant operating cost savings, and is currently being expanded to other sites in North America and Europe.

4 - Optimizing the Orthopedic Supply Chain: Machine Learning-Based Forecasting for Implant Sizing and Surgical Demand in Tka and Tha

Ziqing Zhao, Johnson & Johnson, New Brunswick, NJ, United States

Accurate component sizing in TKA and THA procedures is essential for optimizing clinical outcomes to avoid early failures and excessive costs associated with shipping and maintaining additional surgical trays. Existing methods of implant size predictions for TKA and THA procedures, which include templating, are not very reliable and are time consuming. We have developed and produced machine learning algorithms that leverage real-time patient case data via direct customer Electronic Health & Medical Records (EHR/EMR) integration and provide instant case need predictions. In this work, we developed five ML algorithms that were trained using patient biometrics variables such as height, weight, age, gender, body mass, and surgeon preference indices and productionized the models that exhibited the best accuracies for each procedure. Our models achieve high accuracy rates: up to 95% for ATTUNETM Primary Knee System tibial and femoral components, 80% for hip stems, 87% for hip cups, and 98% for patella; and 80% for the EMPHASYS™ Acetabular Solutions cup. Recently, size prediction models for the EMPHASYS™ cup and ATTUNETM patella were successfully deployed on the Johnson & Johnson MedTech Orthopaedics Advance Case Management platform, a proprietary digitally integrated system that simplifies pre-surgery processes by utilizing case schedules and patient data.

MB17

Building A Level 4 A403

Site Capacity Management and Optimization

Invited Session

Practice Track

Chair: Francisco Castillo Zunino, Amazon, Atlanta, GA, N/A, United States

1 - Site Capacity Optimization: Defining Operational Capacity Frontiers

Francisco Castillo Zunino, Amazon, Luxembourg, Luxembourg, Matias Siebert, Filippo Fedeli, Thomas Øvestad, Inaki Estella Aguerri, Dajun Yue

Accurate capacity planning is a critical challenge in large-scale sorting and distribution centers. While traditional approaches often rely on single-dimensional capacity limits, operational reality shows that different package types compete for shared resources in complex ways. We present a novel approach to site capacity optimization that maps the trade-offs between processing different package types. Rather than defining fixed capacity limits, we develop a mathematical framework that generates multi-dimensional capacity frontiers, revealing how increasing the processing of one package type necessarily impacts the capacity for others. The proposed framework provides facility managers with more precise capacity profiles for operational planning while generating convex constraints suitable for network-wide optimization problems.

2 - Comprehensive Capacity Optimization of the Amazon Transportation Network

Marc Christian Bataillou Almagro, Amazon, Paris, France

Modern transportation networks process millions of packages weekly through integrated planning of labor under-the-roof (UTR) and truck scheduling on-the-road (OTR). This paper introduces a mathematical formulation that bridges virtual package flows with physical transportation constraints, addressing a critical gap in capacity optimization. Our approach integrates operational scheduling constraints with network realities, eliminating infeasible flow plans and reducing capacity adjustments by 10%. We contribute solution methodologies tailored to large scale fulfillment networks: (i) an adaptive discretization scheme that manages high dimensional optimization spaces, (ii) constraint propagation techniques that reduce problem size by 92%, (iii) strategic clustering of commodities and paths that decreases variables by 90%, and (iv) equipment type selection heuristics for optimal truck pre assignment. These innovations enable solving billion variable capacitated network design problems in under 10 minutes.

MB18

Building A Level 4 A404

Decision Analysis in Military Environments

Invited Session

Military and Security

Chair: Tyler Perini, U.S. Naval Academy, Washington, DC, United States

1 - Logistics Resilience in Isolated Environments: Insights from Hawai'i

Isabella Sanders, Dept. of Systems Engineering, United States Military Academy, Newburgh, NY, United States

Isolated regions face outsized disruption risk from hazards, labor shocks, and geopolitical events. We present a simple, transferable framework to assess logistics resilience in isolated environments, illustrated with Hawai'i. Using realistic flow and lead-time estimates across retail supply chains, we simulate targeted lane outages and capacity constraints to quantify unmet demand, recovery time, and stockout propagation. We compare resilience levers, redundancy, calibrated buffers, dynamic rerouting/mode shifts, and information sharing finding two consistent patterns: (1) modest flexibility (transshipment, alternate modes) often outperforms equivalent static redundancy for short, moderate shocks; (2) for sustained disruptions, hybrid strategies that pair right-sized buffers with flexible routing minimize downstream shortages without prohibitive holding costs. The framework supports mission-assured sustainment planning for public and private stakeholders and offers a concise playbook for pre-positioning, diversification, with extensions to cross-sector wargaming and contested-logistics exercises.

2 - Resilience Models for Armament Supply Chains

Benjamin Armstrong, Purdue University, West Lafayette, IN, United States

The resilience of the U.S. military's ammunition supply chain is a critical factor in sustaining combat operations during prolonged or high-intensity conflict. Recent global events and the prospect of near-peer competition have revealed potential vulnerabilities in production, distribution, and surge capacity. This research addresses these challenges by first mapping the ammunition supply chain from the sourcing of raw materials through production and assembly to final distribution and delivery. Developing this comprehensive map provides the foundation for constructing reliable models of current supply chain capabilities.

Drawing on industry-derived data, these models simulate potential surge scenarios to assess vulnerabilities, evaluate scalability, and analyze the cost implications of different courses of action. By quantifying both system performance and cost under stress, the models provide decision makers with a clearer understanding of current readiness and the areas most in need of investment. Beyond modeling, the research also explores alternative materials, manufacturing methods, and readiness-enhancing measures that could improve resilience, provide redundancy in supply and distribution, and enhance flexibility during crises.

This integrated approach offers a framework for strengthening supply chain robustness and adaptability. The findings aim to bridge the gap between industry best practices and military requirements, supporting more informed policy, acquisition, and operational decisions. Ultimately, this work contributes to ensuring ammunition supply remains a reliable enabler of readiness, providing the Army with the capacity to meet the demands of future conflicts without disruption.

3 - Impact of Stochastic Travel TIMES on the Military Port Selection Problem

William Titus, Air Force Institute of Technology, Wright-Patterson AFB, OH, United States, Brian Lunday, Jesse Wales

This research analyzes the impact of stochastic inland and oceanic travel times on port selection for large-scale deployment mobilization. A stochastic mixed-integer programming model seeks to minimize the average arrival time of equipment into theaters. In the first stage, it selects ports to open and assigns installations and ships to open ports. In the second stage, it considers realized travel times and routes equipment to ports and onto sealift ships for overseas transit. Testing applies Sample Average Approximation to a realistic scenario for deployments from the United States to two overseas theaters. Over instances manifesting a range of parametric value combinations, results indicate a critical point of diminishing marginal returns for opening additional ports, as well as relative solution quality for alternative assumptions about the correlation of travel times' stochasticity. Uncorrelated stochasticity reduces port congestion and enables a 2.3% faster average arrival time. For different port opening limitations, subsequent analysis identifies up to a 90% degree of solution similarity, indicating robustness for selected port opening decisions. Testing also identifies that correct assumptions about parametric stochasticity reduce average arrival times by up to 1.2%, indicating the value of validating such information for mobilization planning.

4 - Bayesian Graph Traversal

William Caballero, Air Force Institute of Technology, Wright-Patterson AFB, OH, United States, Phil Jenkins, David Banks, Matthew Robbins

This research considers Bayesian decision-analytic approaches toward the traversal of an uncertain graph. Namely, a traveler progresses over a graph in which rewards are gained upon a node's first visit and costs are incurred for every edge traversal. The traveler knows the graph's adjacency matrix and his starting position, but does not know the rewards and costs. The traveler is a Bayesian who encodes his beliefs about these values using a Gaussian process prior and who seeks to maximize his expected utility over these beliefs. Adopting a decision-analytic perspective, we develop sequential decision-making solution strategies for this coupled information-collection and network-routing problem. We show that the problem is NP-Hard, and derive properties of the optimal walk. These properties provide heuristics for the traveler's problem that balance exploration and exploitation. We provide a practical case study focused on unmanned aerial systems in security operations and empirically study policy performance in myriad Erdos–Renyi settings.

5 - Artificial Intelligence-Enabled Red Team Evaluation of Deep Learning-Powered Military Information Systems

Dinesh Sharma, University of South Florida, Tampa, FL, United States, Ahtesham Bakht, Ankit Shah, Nathaniel Bastian

Military information systems (MIS) collect, process, and analyze critical data to support defense operations and strategic decisions. By integrating complex imagery from spaceborne and airborne hyperspectral sensors, command-and-control (C2) systems for intelligence, surveillance, and reconnaissance (ISR) enhance situational awareness and enable real-time decision-making. While deep learning (DL) boosts the analytical power of MIS, it also exposes them to adversarial machine learning (AML) threats. Understanding these vulnerabilities is essential to advancing the security and robustness of DL-powered MIS. Stealthy poisoning attacks targeting complex imagery remain largely understudied, underscoring the need for robust red teaming frameworks. To address this gap, we introduce the Deep Stealth Poison Generator (DSPG), a novel framework that combines mathematical optimization, deep reinforcement learning, and encoder-decoder architectures to craft highly stealthy and effective poisoning attacks. Experiments with real-world hyperspectral data reveal critical vulnerabilities in DL-powered MIS, with DSPG achieving greater imperceptibility, stronger resistance to data sanitization, and higher targeted misclassification success rates than existing methods. These findings advance the AML literature by deepening our understanding of poisoning threats against DL-powered MIS reliant on complex imagery, and they inform future efforts to improve their robustness and resilience.

MB19

Building A Level 4 A405

Stochastic Systems

Invited Session

Applied Probability Society

Chair: Mark Squillante, Mathematical Sciences Department, IBM Research, Yorktown Heights, NY, 10598, United States

1 - Distributionally Robust Bayesian Control via Optimal Transport: Merton's Problem Revisited

Jose Blanchet, Stanford University, Stanford, CA, United States, Jiayi Cheng, Hao Liu, Yang Liu

Distributionally robust dynamic control often preserves the dynamic programming principle (DPP) for tractability, but this can induce conservatism. We place an optimal transport (OT) ambiguity set on a Bayesian prior for key parameters, relaxing DPP while yielding less pessimistic policies. Our case study is Merton's portfolio selection with medium-horizon objectives (months) in which drift is difficult to estimate, possibly time-varying and unpredictable on these horizons. Volatility, on the other hand, can be screened from higher frequency estimates (hours–days) or implied market estimates. While the screening can affect asset classes, the environment provides a reasonable framework to test the impact of introducing model robustness in a targeted approach. To implement this approach, we develop three results: (i) a min–max duality that collapses the robust control game (without DPP) to a static distributional optimization over priors; (ii) a sensitivity analysis in the ambiguity radius for the resulting nonlinear program, enabling efficient approximations; and (iii) a nonlinear OT projection that calibrates the ambiguity radius from data, generalizing linear projection schemes from statistical distributionally robust optimization (DRO). We evaluate the framework on synthetic and real-market datasets, benchmarking against static DRO and traditional robust control. The experiments clarify when distributionally robust Bayesian control outperforms, and how the gains vary with horizon and volatility screening.

2 - On the (Im)possibility of Beating "Plug-In" Optimization

Henry Lam, Columbia University, New York, NY, United States, Garud Iyengar, Tianyu Wang

Many modern decision-making pipelines entail optimization problems driven by data and the need to mitigate uncertainty. This has ignited a range of proposed approaches in recent years, including various (distributionally) robust optimization and regularized formulations. We explore the fundamental question of when optimization based on simple "plug-in" estimates of unknown quantities is superior among these approaches. Our comparison uses suitable notions of stochastic dominance on the optimality gap or regret, and is based on an uncertainty symmetry concept that holds regardless of whether the uncertainty comes from statistical noise, unpredictable shifts, or both. Finally, we discuss the converse question of when we can improve upon "plug-in" optimizers.

3 - Properties of Turnpike Functions for Discounted Finite Mdp's

Eugene Feinberg, Stony Brook University, Stony Brook, NY, United States, Gaojin He

This paper studies discounted Markov Decision Processes (MDPs) with finite sets of states and actions. Value iteration is one of the major methods for finding optimal policies. For each discount factor, starting from a finite number of iterations, which is called the turnpike integer, value iteration algorithms always generate decision rules which are deterministic optimal policies for the infinite-horizon problems. This fact justifies the rolling horizon approach for computing infinite-horizon optimal policies by conducting a finite number of value iterations. This paper describes properties of turnpike integers and provides their upper bounds.

4 - An Invariance Principle for the Thinned Power of Choice Algorithm

Rami Atar, Technion, Haifa, Israel, Gershon Wolansky

For a thinned version of the power-of-choice, we study an asymptotic regime where the number of servers grows while each queue is kept in "conventional" heavy traffic (i.e., each queue length fluctuates diffusively). Assuming a finite second moment of otherwise general service times we establish an invariance principle, with the diffusion limit given by a McKean-Vlasov SDE.

5 - On a General Optimal Control-Based Approach for Reinforcement Learning

Mark Squillante, Mathematical Sciences Department, IBM Research, Yorktown Heights, NY, United States, Weiqin Chen, Chai Wah Wu, Santiago Paternain

We devise a control-based reinforcement learning approach to support direct learning of the optimal policy, establishing its theoretical convergence, optimality and gradient properties. Numerical results demonstrate significant improvements in quality, complexity, and runtime.

MB20

Building A Level 4 A406

Optimizing Passenger and Freight Rail with Energy and Automation Perspectives.

Invited Session

Railway Applications

Chair: Xuesong Zhou, ITE ASU, Tempe, AZ, United States

1 - Beyond the Horizon: Valuing Flexibility in Urban Rail Extension Decisions Under Political Cycles and Present Bias

Qianwen (Vivian) Guo, Florida State University, Tallahassee, FL, United States, Ziyue Li

This paper formulates and solves the extension problem for urban rail transit lines within a real options framework. Motivated by empirical observations of political cycles inducing quasi-hyperbolic discounting, we incorporate time-inconsistent preferences, specifically present bias, of decision-makers, which violates standard exponential time preference. We employ a quasi-hyperbolic discount function to model preferences within a compound real option valuation framework. The multi-dimensional decision problem involves determining the optimal exercise policy for the compound option, including timing, scale (capacity), and discrete route selection under stochastic project value dynamics. We analyze the strategic implications of time-inconsistency by considering equilibrium concepts for naive and sophisticated agents within an intra-personal game framework. The model derives the optimal exercise boundaries (trigger values) for the investment option. Sensitivity analyses are conducted to quantify the impact of key system parameters and uncertainty levels on the optimal investment policy and resulting project characteristics. This study provides rigorous insights into how time-inconsistency and agent sophistication fundamentally alter optimal infrastructure investment strategies and the value of embedded managerial flexibility.

2 - Intermodal Freight Optimization Through Integrated Railway, Waterway, and Highway Network Modeling

Xuesong Zhou, ASU, Tempe, AZ, AZ, United States

This work enhances intermodal freight logistics by integrating railway, waterway, and highway networks with a focus on operational coordination, disruption resilience, and scalable optimization. Building on POLARIS's regional freight simulation capabilities, we introduce planning and optimization models for intermodal routing, scheduling, and yard operations. Using GMNS+ tools, we construct a multi-resolution freight network that optimizes train blocking plans, block-to-train assignments, and end-to-end shipment flows while accounting for link capacities, yard limits, and transfer times. Our blocking-plan-based traffic assignment enables time-dependent flow simulation and delay resilience analysis under real-world disruptions such as port congestion and demand variability.

3 - Implementation and Analysis of Long-Haul Rail and Intermodal Operations Using Polaris Transportation Framework

Hyunseop Uhm, Argonne National Laboratory, Lemont, IL, United States, Yantao Huang, Abdelrahman Ismael, Olcay Sahin, Hui Shen, Natalia Zuniga, Joshua Auld

The long-haul freight transportation sector faces significant challenges, including a persistent shortage of truck drivers, increasing urban traffic congestion, and instability in the energy supply chain. These factors contribute to significant uncertainty and volatility in shipment costs, particularly for truck-dependent logistics. To address these issues reducing operational reliance on truck-centric operations is a critical strategy for minimizing total transportation costs, improving logistics network stability, and reducing energy consumption.

In this study, POLARIS, an advanced agent-based transportation modeling framework, is used to simulate and analyze intermodal operations. POLARIS is used to generate business-to-business freight shipments, model freight mode and shipment size choices by each commodity flow, simulate freight truck routing and operations, and conduct large-scale transportation simulations across the national long-haul freight network. An optimization model was proposed to minimize total transportation costs by strategically integrating multiple transportation modes (e.g., rail, truck) in intermodal operations.

To support detailed intermodal assessments, the model incorporates train construction and deconstruction, rail schedules, and rail network congestion, enabling more accurate estimations of shipping cost and transit time. The LA-Chicago corridor is used as a case study, pulling from national highway and railway networks. All trucks, railway trains, and shipment containers were dynamically simulated within the POLARIS network.

Finally, the impact on the traffic network and operational delays from this freight redistribution were analyzed under scenarios involving a shift in freight volume from trucks to the railway network.

4 - Train Rescheduling with Considering the Breakdown of the Power Supply System in Metro

Shuai Su, Beijing Jiaotong University, Beijing, China, People's Republic of, Ruihao Zhang

We present a train rescheduling framework for metro systems that accounts for power supply system breakdowns. The methodology optimizes train operations during power failures while minimizing passenger delays and maintaining service continuity. The approach demonstrates improved system recovery and reduced service disruptions.

MB21

Building A Level 4 A407

Re-Inventing Labor Planning at Amazon

Invited Session

Transportation Science and Logistics (TSL)

Chair: Martin Savelsbergh, Amazon Australia, Darlington, WA 6153, Australia

1 - Associate Capacity Forecasting

Selin Tosun, Amazon, Seattle, WA, United States, Jeronimo Callejas

The accurate knowledge on how many Voluntary Extra Time and Ready (on-demand) hours our associate pool has the appetite for in given week becomes crucial in optimal planning and resource allocation, eventually effects the volume processing and efficiency. Hence, many workforce planning modules such as hiring, splitting and planning extra shifts, when and how much head count to account & post, and when & how much to surge is crucial for the operational efficiency. Forecasting the true associate appetite for Voluntary extra time opportunities and Ready opportunities have been challenging, as the historic values has been capped at the head count openings

an extra shift posted with. In other words, the historic data is censored. Similar problems has been seen on retail space, where the real customer appetite trends is unknown when the items go out of stock at certain times. Here in EPS, we study, model, and develop forecasting systems that attacks the censorship in the data by incorporating the censorship within our objective function. We also incorporate the location based embedding coming from census and related datasets. We leverage Fourier Transformations to capture seasonality. We have observed up to 11% improvement with under estimation WMAPE and up to 23% improvement with over estimation WMAPE in the forecasted weekly VET hours, and up to 30% improvement with the under estimation WMAPE and up to 9% improvement with the over estimation WMAPE in the forecasted weekly Ready hours.

2 - Optimizing Capacity: Workforce Regionalization

Hadi Panahi, Amazon, Bellevue, WA, United States, Jeronimo Callejas, Rachel Rutkowski

At the end of Amazon's supply chain, Amazon Logistics (AMZL) delivery stations (DS) must accommodate dynamic customer demand and variations in labor supply reliability, leading to over and under-staffing risk. To minimize this risk, individual DS leverage costly labor planning levers such as Voluntary Extra Time (VET) and surge (i.e., additional compensation). As a cost-effective alternative, we utilized a multi-phased, mixed methods approach to develop and demonstrate the efficacy of a science-based regional labor planning program, referred to as Nodal, which redistributes available labor based on need across nearby stations. We found that Nodal improved labor pool efficiency metrics such as an increase in activity rate of 498 bps (i.e., accepted headcount/eligible headcount), an increase in average hours worked by 430 bps, a decrease in VET utilization by 262 bps, and a decrease in Voluntary Time Off (VTO) utilization by 84 bps. All of these improvements occurred without adversely impacting baseline operational metrics (e.g., attrition, productivity, safety) nor associate or operator satisfaction. These findings suggest that, with guardrails, Nodal can be used to accommodate dynamic customer demand and variation in supply reliability by optimizing the existing AMZL labor pool.

3 - Optimizing Labor use and Availability on the Day of Operations

Nayeon Kim, Amazon, New York, NY, United States, Ramon Auad, Zeynep Sargut

Amazon's commitment to fast delivery options, including same-day service, creates significant demand volatility at facilities until the day of operations. This volatility particularly affects last-mile facilities like Delivery Stations (DS) and Sub-Same Day stations (SSD), requiring agile workforce management to maintain reliable service and lean operations. To address this challenge, Amazon's intra-day planning system dynamically adjusts under-the-roof (UTR) labor plans from the day before operations through the day of operations. The system identifies gaps between existing capacity from current work plans and capacity needs based on updated forecast. It leverages Voluntary Extra Time (VET) and Ready labor to address understaffing and Voluntary Time Off (VTO) to minimize labor costs during periods with overstaffing. The system also optimizes workforce assignment across different processes throughout the day to maximize labor utilization and total throughput. In this presentation, we outline the system architecture and its supporting modules that collectively enable flexible labor planning while maintaining operational reliability and cost efficiency. Each module addresses a specific challenge as a generic solution, such as optimal VTO planning. This modular approach achieves better scalability and explainability while allowing adaptation to different planning horizons and business contexts, ensuring broad applicability across Amazon operations.

4 - Conceptualizing Labor Planning as a Dynamic Process: Developing a Versatile Qualitative Dataset to Support Large Language Models

Alana Scholl, Amazon, New York, NY, United States

Labor planning for Under-The-Roof (UTR) operations is a complex process characterized by numerous inputs, reliance on multiple roles for feedback and approval, site-specific nuances, time-sensitivity, and adverse outcomes for the business, operators, and associates should labor plans fail to reflect reality. In an effort to optimize the time and effort of labor planners, there have been efforts to develop real-time support resources, such as large language model chatbots, to automate and expedite aspects of the labor planning process. However, to do this effectively, we must have a comprehensive understanding of *what* planners do, how they execute their tasks, and how they make decisions while navigating disparate tradeoffs. Existing standard operating procedures (SOPs) do not capture this level of detail. As such, we leveraged a modified hierarchical task analysis (HTA) approach using three forms of data collection including 1) artifact analysis of existing SOPs and subject matter expert interviews, 2) structured observations and modified contextual-inquiry based observations the labor planning process, and 3) modified critical incident technique interviews with planners. Using this approach, we were able to effectively and efficiently capture the labor planning process as a series of contextualized, interdependent decision points with an HTA structure and process map as initial outputs. From there, we were able to reformat the data to support multiple use cases including: LLM fine-tuning, updating onboarding materials, and developing SOPs to establish and monitor quality standards.

5 - Large Language Models for Model Explainability: a use-Case in Labor Planning

Anis Ben Said, Amazon, Seattle, WA, United States, Roman Levkin, Thomas Fillebeen, David Dong, Brian Salisbury

Amazon's Under-The-Roof (UTR) operations—comprising fulfillment centers, sort centers, and delivery stations—employ hundreds of thousands of associates and require sophisticated labor planning. Current labor planning models provide complex planning decisions by balancing demand fulfillment and labor cost minimization, while accounting for numerous operational constraints, influencing hiring and scheduling decisions for hundreds thousands associates. However, these models often act as “black boxes” for the planners who use them, who do not have specialized optimization knowledge. As a result, planners spend large amount of time investigating model recommendations and tweaking model inputs to derive their desired results, potentially leading to reduced trust and suboptimal manual adjustments.<div c-

Recent advancements in Large Language Models (LLMs) have demonstrated capabilities in explaining complex systems such as code bases and mathematical models. This research introduces a generalizable framework for model explainability through LLM agents and applies it to the labor planning space. The framework presents a modular architecture design alongside implementation methods that enable rapid development of GenAI based chatbots specialized in explaining diverse model types. While our initial focus has been on optimization models for labor planning, the approach can be adapted to any model requiring human interpretation. Using this framework, we have successfully onboarded multiple labor planning models, demonstrating significant improvements in planner efficiency and decision quality.

MB22

Building A Level 4 A408

APS Student Paper Competition

Award Session

Applied Probability Society

Chair: David Goldberg, Cornell University, Ithaca, NY, United States

1 - Optimizing LLM Inference: Fluid-Guided Online Scheduling with Memory Constraints

Ruicheng Ao, MIT, Cambridge, MA, United States

Large Language Models (LLMs) are indispensable in today's applications, but their inference procedure requires substantial computational resources, especially when memory is limited. This paper treats LLM inference optimization as a multi-stage online scheduling problem, where prompts arrive sequentially and the incremental expansion of the Key-Value (KV) cache during inference renders conventional scheduling algorithms ineffective. To address this challenge, we develop a fluid dynamics approximation to establish a tractable benchmark, providing insights for devising effective scheduling algorithms. We introduce the Waiting for Accumulated Inference Threshold (WAIT) algorithm for known output lengths and extend it to the Nested WAIT algorithm for unknown output lengths. Both algorithms achieve near-optimal performance compared with the fluid benchmark under heavy traffic limits, balancing throughput, latency, and Time to First Token (TTFT). Numerical experiments with Llama-7B on an A100 GPU demonstrate superior throughput and reduced latency compared to widely adopted baselines.

2 - Efficient Gradient Estimation for Overparameterized Stochastic Differential Equations

Shengbo Wang, University of Southern California, Los Angeles, CA, United States

Overparameterized stochastic differential equation (SDE) models have achieved remarkable success in various complex environments, such as PDE-constrained optimization, stochastic control and reinforcement learning, financial engineering, neural SDEs, and generative AI. These models often feature system evolution coefficients that are parameterized by a high-dimensional vector θ in $\mathbb{R}^n$, aiming to optimize expectations of the SDE, such as a value function, through stochastic gradient ascent. Consequently, designing efficient gradient estimators for which the computational complexity scales well with n is of significant interest. We introduce an unbiased stochastic gradient estimator—the generator gradient estimator—for which the computation time remains stable in n . In addition to establishing the validity of our

methodology for general SDEs with jumps, we also perform numerical experiments testing our estimator in a linear quadratic control problem where the policy is parameterized by high-dimensional neural networks. The results show a significant improvement in efficiency compared to previous methods: our estimator achieves near-constant computation times, increasingly outperforms its counterparts as n increases. These empirical findings highlight the potential of our proposed methodology for optimizing SDEs in contemporary applications.

3 - Vc Theory for Inventory Policies

Yaqi Xie, University of Chicago, Chicago, IL, United States

There has been growing interest in applying reinforcement learning to inventory management, alongside a recent stream of work advocating for supervised learning approaches that directly optimize inventory decisions using full historical demand trajectories. While empirically promising, there has been no theoretical understanding of supervised learning in inventory management problems, where inventory is dynamically replenished over time. This paper provides the first generalization guarantees for supervised learning of several well-known classes of dynamic inventory policies, including base-stock and (s, S) policies, using tools from the celebrated Vapnik-Chervonenkis (VC) theory, such as the Pseudo-dimension and Fat-shattering dimension. Our results have implications for regret against the best-in-class policies, and allow for an arbitrary distribution over demand sequences, which makes no assumptions such as independence across time. Surprisingly, we prove that the class of policies defined by TS non-stationary base-stock levels exhibits a generalization error that does not grow with TS , whereas the two-parameter (s, S) policy class has a generalization error growing logarithmically with TS . Overall, our analysis leverages specific inventory structures within the learning theory framework, and improves sample complexity guarantees even compared to existing results assuming independent demands.

4 - Distributed Load Balancing with Workload Dependent Service Rates

Wenxin Zhang, Columbia University, New York, NY, United States

We study distributed load balancing in bipartite queueing systems where frontends route jobs to heterogeneous backends with workload-dependent service rates. The system's connectivity -- governed by compatibility constraints such as data residency or resource requirements -- is represented by an arbitrary bipartite graph. Each frontend operates independently without communication with other frontends, and the goal is to minimize the expected average latency of all jobs. We propose a closed-loop policy called the Greatest Marginal Service Rate (GMSR) policy that achieves effective coordination without requiring knowledge of arrival rates. In a discrete-time stochastic model, we show that the behavior of our routing policy converges (almost surely) to the behavior of a fluid model, in the limit as job sizes tend to zero and job arrival rates are scaled so that the expected total volume of jobs arriving per unit time remains fixed. Then, in the fluid regime, we demonstrate that the policy attains an ϵ -suboptimal solution in $O(\delta + \log 1/\epsilon)$ time from δ -suboptimal initial workloads, which implies global convergence to the centrally coordinated optimal routing. Finally, we analyze the fluid model when the system is overloaded. We show that GMSR lexicographically maximizes throughput, maximizes the number of stable backends, and minimizes their collective workload.

MB23

Building A Level 4 A409

Panel with INFORMS Editors-in-Chief on Writing Strategies

Panel Session

Minority Issues Forum

Co-Chair: Sibel Alumur Alev, University of Waterloo, Waterloo, ON, Canada

1 - Moderator Panelist

Sibel Alumur Alev, University of Waterloo, Waterloo, ON, Canada

2 - Panelist

Oktay Gunluk, Georgia Tech, Atlanta, GA, United States

3 - Panelist

Guillaume Roels, INSEAD, Fontainebleau, France

4 - Panelist

Karen Smilowitz, Northwestern University, Evanston, IL, United States

5 - Panelist

Amy Ward, The University of Chicago Booth School of Business, Chicago, IL, United States

6 - Panelist

Alice Smith, Auburn University, Auburn, AL, United States

MB24

Building A Level 4 A410

Mitigating the Impacts of Wildfires on Electric Power Systems through Stochastic Optimization

Tutorial session

Tutorials

Chair: Brian Denton, University of Michigan, Ann Arbor, MI, United States

1 - Mitigating the Impacts of Wildfires on Electric Power Systems Through Stochastic Optimization

Chaoyue Zhao, University of Washington, Seattle, WA, United States, Ruiwei Jiang, Juan-Alberto Estrada-Garcia, Xinyi Zhao, Alexandre Moreira

Dry and windy weather conditions significantly increase the risk of wildfires, whose spread exacerbates the vulnerability of the grid and results in prolonged power outages. This tutorial introduces and reviews recent streams of studies on addressing this challenge through stochastic optimization approaches, including static, adaptive, dynamic, and distributionally robust models. In particular, we account for random failures of power lines, which depend not only on the ambient environment (such as temperature, wind speed, and fire) but also on the power flowing through the line, introducing decision-dependent uncertainty (DDU). We introduce the modeling of wildfire, power systems operations, and their interactions, as well as how stochastic optimization models can characterize DDU and mitigate the impacts of wildfires on electric power systems. As examples, we mention three models, ranging from long-term planning to short-term and dynamic reconfiguration of a power system amidst wildfire-prone conditions. For each model, we provide a numerical case study to demonstrate the value of modeling (e.g., DDU and dynamic reconfiguration) in mitigating the impacts of wildfires.

MB25

Building A Level 4 A411

Uncertainty in Resource Allocation

Invited Session

TSL: Freight Transportation

Chair: Christine Nguyen, Northern Illinois University, Dekalb, IL, United States

1 - Technician Routing/Scheduling and Parts Management with Demand Uncertainty**Stephanie Santiago-Montano, Auburn University, Auburn, AL, United States, Daniel Silva, Alice Smith**

The technician routing and scheduling problem is encountered in several industries where equipment installation/maintenance at different locations is required. As factors such as demand uncertainty, time windows, skilled staff requirements, and multiple visits come into play, the problem complexity increases. Moreover, for home service companies (HVAC, plumbing, electrical, etc.) to maintain a competitive advantage, they need efficient management of their limited resources. Technological advances like drones can help make this part of the supply chain more cost-effective, faster, and eco-friendly. The combination of traditional trucks and drones can provide an efficient system in which vehicles complement each other to overcome the challenges/disadvantages that each one has. The efficiency of home service networks greatly depends on proper inventory management and efficient technician routing/scheduling. Most studies focus on these problems individually. Considering both simultaneously could result in cost savings and an overall network efficiency increase. This work studies an integrated routing/scheduling and inventory management problem with stochastic demands using a truck/drone tandem. In this problem, demands for spare parts are revealed throughout the day (i.e., demands are placed during a technician visit to the customer location, when the technician diagnoses the needs) and are unknown before the request time. To solve this problem, a rolling-horizon algorithm that seeks the best possible myopic policy is developed. Instances are generated to compare the performance of traditional truck-based networks and the proposed tandem network. Furthermore, this approach is compared to an alternative that waits until multiple demands are revealed. Managerial insights are then provided.

2 - Network Restoration Problem with Heterogeneous Crews Under Incomplete Information**Youssef Nsouli, Rensselaer Polytechnic Institute, Troy, NY, United States, Bahar Cavdar**

We consider a disrupted infrastructure network in the aftermath of a disaster where the planners know the locations of the service outages, but have incomplete information on the actual damage locations. The planner's goal is to determine the routes for two heterogeneous repair crews, a high-skill and a low-skill crew, where the low-skill crew can only repair a subset of damages, to minimize the total service disruption duration experienced by all customers. We formulate this problem as a Markov decision process. We first analyze the structural properties of optimal policies; we consider the one-crew (high-skill) only variant of the same problem and explore how to integrate the low-skill crew without impeding the high-skill crew. We then develop a solution method based value approximation using reinforcement learning, integrating our structural results.

3 - Data-Driven Bunker Refueling Under Price Uncertainty**Qinghe Sun, The Hong Kong Polytechnic University, Hong Kong, Hong Kong**

Bunker refueling decisions in international shipping are crucial operational choices. Each ship acts like a movable storage unit navigating through diverse markets, procuring bunker fuels from different ports to sustain its voyage. This involves grappling with challenges posed by varying bunker fuel prices over time and locations. To address this, we propose data-driven structure-prescriptive (SP) approaches that combine the strengths of modern machine learning with the insights from traditional OR modeling and optimization. Instead of predicting future fuel prices, our approach directly learns the optimal policy from data and adapts refueling decisions to the current market conditions, including fuel prices, crude oil price, NYSE index, etc. Our focus lies in leveraging the well-established understanding that the optimal refueling decision adheres to a state-dependent base-stock refueling policy. This decision depends on factors such as the port of call, fuel tank capacity, market conditions, and is finite-valued, depending on the ship's schedule and voyage. We provide a practical framework to incorporate these structural properties into data-driven decision-making. The finite-valued property, for instance, establishes a connection between the refueling problem and multiclass classification, integrating OR modeling with machine learning. We test our approach through experiments under both simulated scenarios and a real-world example involving the Asia-North America Trade Route. The proposed SP approaches successfully recovered the "true" optimal refueling policy in synthetic simulations. Our real-world data experiments further demonstrate that integrating structural properties into the learning process enhances the quality of prescriptions, particularly when training data is limited.

4 - How to use the Concept of Fragility to Guide the Design of Robust Timetables**Marcella Samà, Roma Tre University, Rome, Italy, Marta Leonina Tessitore, Giorgio Sartor, Carlo Mannino, Dario Pacciarelli**

A robust timetable is crucial for any efficient rail transport system. A typical way to increase the robustness of a timetable is to include buffer times, which can be used to mitigate delays. Other approaches, instead, improve the timetable robustness by considering the possibility of mitigating delays in real-time through dispatching. In this work, we propose a Mixed Integer Linear Programming model to build robust timetables by focusing on their most vulnerable components. This model comprises a set of timetable variables and associated constraints, and a collection of dispatching sub-models related to the fragility of the considered scenarios. We apply our methodology to a real-world case study involving a Norwegian railway line to demonstrate the significant benefits of applying a fragility-based approach to timetable design. We start by computing the fragility of a timetable, for each train-resource pair, by solving a dispatching problem for each delay scenario. Focusing on the most fragile train-resource pair, we apply a delayed row generation scheme in order to dynamically generate most timetabling and dispatching conflict avoiding constraints in the model. We then integrate the proposed algorithm into an iterative process, where the timetable is repeatedly optimized to reduce the fragility of the most fragile train-resource pair. By doing this, we are able to reduce the fragility of the most critical pair, in each iteration, even when keeping the new timetable as close as possible to the original one.

5 - Stochastic Supply Chain Network Design for Fair Resource Allocation**Christine Nguyen, Northern Illinois University, Dekalb, IL, United States, Gabriela Ibarra, Mike Hewitt**

Acting as a vital link between food-insecure individuals and locally sourced food donors, not-for-profit food distributors provide millions of meals to individuals in need. Not-for-profit food distribution networks typically consist of two key members: donors and food banks. Donors provide physical goods and food banks allocate the donations. Effective collaboration between donors and food banks plays a vital role in reducing food waste, and addressing a major concern for these networks – the fair allocation of food. Additionally, the uncertainty of supplier donations and the broad range of variability of demand over time can significantly impact the efficiency of the network. By incorporating the probability of these uncertainties, a network that is more adaptable and responsive to fluctuations in supply and demand can be developed. A two-stage stochastic program is proposed to design a not-for-profit food distribution supply chain network that integrates demand and supply uncertainty, ensures customer satisfaction through service levels, and provides fair allocation. To better serve customers, it is important to consider several strategic and operational questions related to the food distribution network. Research questions in the first stage include: (i) how many consolidation centers should be established and where should they be located? (ii) how many vehicles should be owned? In the second stage: (i) how would external transportation services – such as third-party logistics– supplement the delivery routes from owned vehicles? Disaster relief efforts, blood supply management, and organ transplant operations can greatly benefit from the proposed research.

MB26

Building A Level 4 A412

SOLA Award

Award Session

Location Analysis

Chair: Karmel Shehadeh, University of Southern California, Los Angeles, CA, United States

1 - Spin the Bottle Bill: Deposit-Refund System Policy and Reverse Supply Chain Designs**Austin Saragih, Massachusetts Institute of Technology, Cambridge, MA, United States**

The Bottle Bill, a deposit-refund system policy, charges consumers a deposit upon beverage purchase and refunds it upon return of the empty container to a recycling center. In California and nine other states, the redemption value has remained stagnant at 5 cents for decades, the number of recycling centers has decreased, and recycling

rates have been declining. To address these issues, we spin the Bottle Bill by jointly optimizing its policy and reverse supply chain as a bilevel mixed-integer nonlinear program. At the policy design level, we optimize the redemption value, manufacturers' fees, and distance policy between consumers and recycling centers. Given this distance policy, at the network design level, we then optimize the number of recycling centers for each county in the state. To solve the bilevel problem, we devise a novel solution approach using continuous approximation and monotonicity analysis with optimality and convergence guarantees. Applying our approach to California's Bottle Bill, we provide policy options and insights that the state could implement to achieve its 80% recycling goal, by explicitly and jointly balancing its redemption value and the number of recycling centers. These options and insights can also be adapted to other U.S. states and countries to advance the circular economy of containers.

2 - Effects of Geopolitical Strain on Global Pharmaceutical Supply Chain Design and Drug Shortages

Martha L. Sabogal De La Pava, Clemson University, Clemson, SC, United States, Emily Tucker

Emerging geopolitical risks have begun to threaten global supply chains, including those that produce life-saving drugs. Export bans may prevent a company from shipping products internationally, and it is unclear how these new dynamics may affect company plans and persistent, worldwide drug shortages. To address these questions, we present a global pharmaceutical supply chain design model that considers the risk of export bans that are induced by supplier capacity disruptions and corresponding price increases. The model takes the company's perspective as a decision-maker looking to locate plants and distribute drugs globally. It is a two-stage stochastic program that includes uncertainty in capacity, ability-to-export, and demand. The model is solved by integrating the Sample Average Approximation and L-shaped methods. We present conditions related to when demand will be met and a case study of a generic oncology drug. We find that preparing for geopolitical strain may increase resilience and profits as well as reduce shortages in the short term. At baseline, expected global shortages are high (17.2%) with disparities across country income levels (0.3%, 0.8%, 87.2%, and 87.6% for high, upper-middle, lower-middle, and low income countries, respectively). Pricing policies may improve drug access overall, back-shoring may slightly improve access for the country where it is implemented, and bilateral alliances may not be effective at improving access.

3 - When Where Watt: Harnessing the Value of Time and Location for Renewable Electricity Generation

Vishrut Rana, The Wharton School, Philadelphia, PA, United States, Christian Kaps, Serguei Netessine

Global investment in renewable energy has surpassed \$2 trillion, propelled by cheaper technologies, supportive policies, and energy security needs. Yet profits from renewable generation remain modest, well below the returns from fossil fuel plants, because renewable electricity from wind and solar is often generated at times and in places with low wholesale prices. These lower returns are a central barrier to scaling renewables in line with global climate goals. To enhance profitability of renewable projects, we introduce *Quality Adjusted Power Value (QAPV)*, a siting metric that incorporates the time and location of generation and local prices into investment decisions. Using geospatial satellite data and electricity prices from Texas, we find that 59 wind plants built between 2008 and 2014 would earn median 21% higher annual revenue if sited using QAPV. These gains persist even after accounting for political and infrastructure constraints. We also estimate that marginal wind generation lowers local market prices by 3.8%, translating to consumer benefits of \$1.53 million per site-year under QAPV siting. QAPV thus enables developers to prioritize higher-revenue sites for investment and helps policymakers design support schemes that reward quality rather than just quantity of renewable generation.

4 - Identifying Regions Vulnerable to Obstetric Unit Closures Using Facility Location Modeling with Patient Behavior

Meghan Meredith, NYU Langone Health, Atlanta, GA, United States

The United States has the highest maternal mortality rate among high-income countries. This is just the tip of the iceberg, with many more women experiencing severe complications due to their pregnancy. One commonly cited challenge in maternal healthcare delivery is that many pregnant people lack access to obstetric care, and this has been exacerbated by widespread closures of obstetric units. In this talk, we present a framework for identifying regions that are vulnerable to obstetric unit closures using facility location modeling. When considering the impacts of closures, we incorporate patient care-seeking behavior as many pregnant people "bypass" their closest obstetric facility to deliver at a farther facility. Altogether, we consider a problem in which an adversary seeks to maximize a population's average distance traveled to obstetric care by closing obstetric units. After the adversary closes obstetric units, patients probabilistically select the facility that maximizes their utility under a random utility maximization framework. We demonstrate that this model can be represented as a mixed-integer nonlinear program. Unlike existing facility location models that incorporate consumer choice, our model's objective function is not concave in general and thus, the state-of-the-art algorithms for solving facility location models with choice are not applicable. We introduce reformulations of our problem that lend themselves to solution methods that can take advantage of their decomposable structure. We compare our solution methods in numerical experiments and illustrate our framework using a case study of obstetric care in the state of Georgia.

MB27

Building B Level 2 B201

Discrete Bilevel Optimization

Invited Session

Computing Society

Chair: Sebastián Vázquez, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, United States

1 - Game-Theoretic Relaxations in Bilevel Optimization

Noah Weninger, University of Waterloo, Waterloo, ON, Canada, Ricardo Fukasawa

We introduce a framework for relaxing Stackelberg zero-sum games with perfect information, with a focus on those arising from discrete bilevel optimization. The framework consists of operators which transform games while maintaining an approximation or bound on the optimal objective value. This simplifies and generalizes our two previous works on knapsack and matroid interdiction. It also generalizes well-known techniques including the high point relaxation and the derivation of bounds by restricting/relaxing the lower level problem. To further motivate this approach, we apply it to matching interdiction, in which the objective is to interdict (i.e., remove) edges from a graph so that the maximum matching weight is minimized, subject to a knapsack constraint on the set of interdicted edges. We present a branch-and-bound algorithm for this problem which uses a new lower bound derived using our framework.

2 - Solving Integer Robust Optimization Problems with Integer Decision-Dependent Uncertainty Sets

Leonardo Lozano, University of Cincinnati, Cincinnati, OH, United States, Juan Borrero

We study a class of decision-making problems under uncertainty, where the planner hedges against the worst-case realization in an integer uncertainty set that has a functional dependence on the decisions of the planner. We propose three methods to reformulate and solve these problems: the first is based on a single-level 'semi-infinite' reformulation, where the dependence of the uncertainty on the decisions is modeled with auxiliary binary variables; the second is based on a single-follower bilevel reformulation; and the third one is a multiple-follower bilevel reformulation where each robust constraint is associated with a follower. We provide additional enhancements to strengthen the linear programming relaxation of the single-level formulation and propose a column and constraint generation algorithm to solve it. The multiple-follower bilevel formulation is further transformed into a multicommodity flow single-level mixed-integer problem by exploiting recent advancements in decision diagrams. We perform numerical experiments that show the column and constraint generation algorithm can solve small- and medium- scale instances within a few minutes. This method is outperformed by the multicommodity flow formulation when dealing with problems with a single robust constraint. We complement our experiments by using the column and constraint generation algorithm to solve a constrained shortest path problem with maintenance decisions and uncertain arc failures. The proposed method can solve instances of this problem with hundreds of nodes and thousand of arcs in a few minutes.

3 - A Bilevel Optimization Approach for A Class of Combinatorial Problems with Correlated Disruptions and Probing

Jaime Y. Perez, University of South Florida, Tampa, FL, United States, Juan Borrero

We study linear combinatorial optimization problems in which uncertain disruptions can affect the objective function coefficients. A decision maker, or planner, can allocate a limited probing budget to fully reveal the disruption status of selected components. Then, in a second stage, the planner solves the combinatorial problem and executes a solution plan based on the partial information discovered by probing and estimating the remaining uncertainty. In contrast to traditional settings assuming independence, we consider statistically dependent disruptions, allowing probabilities updates about non-probed components. After executing a solution plan, the

uncertainty realizes and the true reward is computed. The goal is to select a probing plan that maximizes the expected true reward. We reformulate the probing problem as a stochastic mixed-integer nonlinear bilevel problem with multiple followers and propose two solution approaches. The first uses the optimal value function reformulation, where conditional probabilities are expressed by a ratio of multilinear polynomials. This yields a large-scale problem with an exponential number of variables, which we address via a cutting-plane algorithm. The second approach applies a strong-duality reformulation when the second-stage problem is a linear program, representing conditional probabilities as maximum flow problems over a graph constructed by the scenario set of uncertainties. A specialized branch-and-bound algorithm exploits the resulting reformulation structure by solving the problem for each scenario in parallel. Computational experiments for the longest path problem show that probing a subset of the components can yield large improvements in the optimal value for small- to medium size instances.

4 - Decision Diagram-Based Cuts for Binary Bilevel Programs

Sebastián Vásquez, Carnegie Mellon University, Pittsburgh, PA, United States, Leonardo Lozano, Willem-Jan van Hoeve

Solving binary bilevel programs remains challenging due to the lack of strong, tractable relaxations. Recent work has embedded decision diagrams into the high-point relaxation to obtain exact extended reformulations and valid bounds. However, diagram size often limits scalability. We propose a new decision diagram-based framework for binary bilevel programs that uses diagrams as oracles to derive cuts directly in the original variable space. Building on existing approaches, we introduce new exact reformulations and show how to improve structured cuts. Moreover, we introduce a new extended diagram encoding that allows to retrieve stronger cuts and improve any valid cut. Computational results demonstrate the effectiveness of our approach.

MB28

Building B Level 2 B202

LLMs and Optimization

Invited Session

Computing Society

Chair: Connor Lawless, Stanford University, San Francisco, United States

1 - An Update to Optichat: Explaining Optimization Models Using Large Language Models

Can Li, Purdue University, West Lafayette, IN, United States, Hao Chen, Sai Madhukiran Kompalli

A significant obstacle to the practical deployment of optimization models is the challenge associated with helping practitioners comprehend and interpret these models. Our research group has been developing OptiChat, a chatbot powered by large language models, designed to explain optimization problems. The initial version of OptiChat supports a variety of functionalities, including diagnosing infeasibilities, conducting sensitivity analyses, offering counterfactual explanations, and answering general user queries. In this talk, we present recent updates to OptiChat, including integration with multiple LLM platforms, incorporation of interpretable decision rules, support for more complex counterfactual queries, and the generation of graph-based sensitivity insights. These enhancements aim to further bridge the gap between optimization model outputs and user comprehension.

2 - Llms for Cold-Start Cutting Plane Separator Configuration

Anders Wikum, Stanford University, Stanford, CA, United States, Connor Lawless, Yingxi Li, Madeleine Udell, Ellen Vitercik

Mixed integer linear programming (MILP) solvers ship with a staggering number of parameters that can have an outsized impact on the performance of the MILP solver, but are challenging to configure for all but expert optimization users. Existing machine learning (ML) approaches for solver configuration require solving thousands of related MILP instances during training, generalize poorly to new problem sizes, and often require implementing complex ML pipelines and custom solver interfaces that can be difficult to integrate into existing optimization workflows. In this paper, we introduce a new LLM-based framework to configure which cutting plane separators to activate for a given MILP problem using little to no training data. Our approach produces separator configurations based on characteristics of the instance, like a LaTeX formulation and natural language description, and is augmented by literature-based descriptions of the cutting plane separators available in a given solver. While individual configurations have a large variance in performance, we present a novel ensembling strategy that clusters and aggregates configurations to create a small portfolio of high-performing configurations. Our LLM-based methodology requires no custom solver interface, generates individual configurations with simple API calls that run in under a second, and can find a high-performing configuration by solving only a small number of MILPs. Numerical results show our approach is competitive with existing configuration approaches on a suite of classic combinatorial optimization problems and real-world datasets, using a fraction of the training data and computation time.

3 - Feasibility Versus Optimality in the Verification of Neural Networks Using Mixed Integer Linear Programming

Konstantinos Ziliaskopoulos, Auburn University, Auburn, AL, United States, Alice Smith

The formal verification of neural network properties presents significant computational challenges due to the non-convex nature of activation functions, particularly the Rectified Linear Unit (ReLU), which introduces binary decisions into the optimization model. Mixed-Integer Linear Programming (MILP) formulations offer a natural framework for modeling these networks by representing ReLU functions as complementary constraints with binary indicators, enabling rigorous verification guarantees and adversarial sample discovery. However, formulating neural networks as mathematical models comes with the cost of computational complexity that grows exponentially with network size. This research examines alternative MILP formulation approaches for neural networks. We explore the transformation of traditional optimization problems with distance metric objectives (e.g., L-infinity, L1 norms) into feasibility problems that search for adversarial examples within specified error tolerance bounds. Our methodology compares these approaches across multiple dimensions, including solution times, preprocessing effectiveness, objective formulation, and constraint structure. Additionally, we investigate pre-solving techniques for variable bound tightening and methods for eliminating redundant piecewise linear constraints before optimization. The study evaluates the interpretability of various objective function metrics and their use cases and examines misclassification rates after retraining models with adversarial samples. Through systematic comparison of these formulation strategies, this work contributes to the understanding of computational trade-offs between MILP formulation complexity, solution approach, and adversarial training frameworks in neural network verification.

4 - Equivamap: Leveraging LLMs for Automatic Equivalence Checking of Optimization Formulations

Haotian Zhai, The University of Texas at Austin, Austin, TX, United States, Connor Lawless, Ellen Vitercik, Leqi Liu

A fundamental problem in combinatorial optimization is identifying equivalent formulations. Despite the growing need for automated equivalence checks—driven, for example, by *optimization copilots*, which generate problem formulations from natural language descriptions—current approaches rely on simple heuristics that fail to reliably check formulation equivalence. Inspired by Karp reductions, in this work we introduce *Quasi-Karp equivalence*, a formal criterion for determining when two optimization formulations are equivalent based on the existence of a mapping between their decision variables. We propose *Equivamap*, a framework that leverages large language models to automatically discover such mappings for scalable, reliable equivalence checking, with a verification stage that ensures mapped solutions preserve feasibility and optimality without additional solver calls. To evaluate our approach, we construct *Equivamap*, the first open-source dataset of equivalent optimization formulations, generated by applying transformations such as adding slack variables or valid inequalities to existing formulations. Empirically, *Equivamap* significantly outperforms existing methods, achieving substantial improvements in correctly identifying formulation equivalence.

5 - Heuristics for a Vehicle Routing Problem with Size Constraints in Agricultural Applications, Including An LLM-Based Approach

Jung Yun Bae, Michigan Technological University, Houghton, MI, United States, Hyeseon Lee

Vehicle routing is a critical challenge in coordinating heterogeneous multi-robot systems for agricultural applications. This research addresses a vehicle routing problem in automating lavender harvesting, where robots of varying sizes must collaboratively navigate complex field layouts to efficiently complete given tasks.

We propose two heuristic approaches to address a specific problem of vehicle routing for lavender harvesting. The first approach utilizes a Large Language Model (LLM) to generate harvesting plans. Providing the LLM with small-scale examples and iteratively refining prompts with detailed descriptions of task attributes, robot capabilities,

and environmental conditions produces feasible task allocations and paths for each robot while minimizing overall operational time. The second approach employs a greedy heuristic algorithm, which starts with an initial feasible solution and iteratively improves it by optimizing task allocation and robot paths while ensuring all constraints are satisfied. This method guarantees feasibility by directly incorporating task requirements and robot constraints into its optimization process.

Both approaches are validated through simulations of lavender fields with varying robot sizes, environmental layouts, and numbers of robots and tasks. The results demonstrate the effectiveness of these methods in achieving efficient routing for heterogeneous multi-robot systems. This work highlights the potential for these approaches to advance automation in lavender harvesting, contributing to increased efficiency and sustainability in agricultural operations, further considering fuel and payload constraints as well as required harvesting techniques.

MB29

Building B Level 2 B203

Hotelling Medal for Lifetime Achievement in ENRE

Award Session

Energy, Natural Resources and the Environment

Chair: Gokce Palak, Wake Forest University, Winston-Salem, NC, United States

Co-Chair: Benjamin Leibowicz, The University of Texas at Austin, Austin, TX, United States

1 - Introducer - Chair

Shmuel Oren, UC Berkeley, Berkeley, CA, United States

Introducer - Chair

2 - An Old Theorem and A Chaotic World

Yves Smeers, Université Catholique de Louvain, Mont-Saint-Guibert, Belgium

Based on “Marginal Pricing and the Energy Crisis: where should we go?” I. Abada, A. Ehrenmann and Y. Smeers; forthcoming in “Energy Economics”. Abstract: The old theorem is due to M. Boiteux (1960). It states the equality between short and long-term marginal costs when capacities are adapted. The chaotic world is the one where shocks prevent capacities to be adapted, which seems to become the new standard. The question is what remains from Boiteux’s theorem when his fundamental assumption of adapted capacities is no longer satisfied and what are the implications for capacity development and capacity expansion models. These ideas developed from the energy crisis in Europe after the eruption of the war in Ukraine; they may remain relevant today and extend beyond the power sector.

MB30

Building B Level 2 B204

Machine learning as a Tool in Power and Energy

Invited Session

ENRE: Electricity

Chair: Robert Mieth, Rutgers University, Piscataway, NJ, United States

1 - Data-Driven Successive Linearization for Optimal Voltage Control

Wenqi Cui, New York University, New York City, NY, United States, Steven Low

High variability of solar PV and sudden changes in load (e.g., electric vehicles and storage) can lead to large voltage fluctuations. In recent years, a number of advanced controllers have been designed to optimize voltage control. However, these controllers typically rely on fixed linear approximations of voltage dynamics. As a result, the solutions may become infeasible when applied to the actual voltage behavior governed by nonlinear power flow equations, particularly under heavy power injection from distributed energy resources. This talk will describe a data-driven successive linearization approach for voltage control under nonlinear power flow constraints. By leveraging the fact that the deviation between the nonlinear power flow solution and its linearization is bounded by the distance of the operating point, we perform data-driven linearization around the most recent operating point. Convergence of the proposed method to KKT points is established by exploiting the convexity of the objective function and the structural properties of the nonlinear constraints. Case studies show that the proposed approach achieves fast convergence and adapts quickly to changes in net load.

2 - Data as a Requisite for Learning in Power Systems

Rabab Haider, University of Michigan, Ann Arbor, MI, United States

Machine learning models have been developed for a wide variety of power system applications. In many settings where real data is limited or unavailable, machine learning models are trained using synthetic data sampled via different strategies. Our preliminary results show that the methodologies employed to generate these datasets greatly influence learning efficacy. This result is not surprising. Specific to power systems, different data streams, including topological data, generator and load characteristics, and voltage measurements impact the ability for neural networks to learn the underlying power flow representation. Data augmentation approaches, including active and importance sampling, are well established in literature, but may be limited if the data streams themselves are limited. Recommendations for dataset generation and augmentation may be derived, and additional work to develop datasets for learning in power systems is needed.

3 - Optimistic Bias in Data-Driven Risk Estimation and Optimization

Utsav Sadana, Université de Montreal, Montreal, QC, Canada, Erick Delage, Angelos Georghiou

The entropic risk measure is commonly used in high-stakes decision-making to account for tail risks. Empirical entropic risk estimator that replaces expectation in the entropic risk measure with sample average underestimates true risk. To correct this bias, a strongly asymptotically consistent bootstrapping procedure is proposed that fits a distribution to the data and then estimates the bias in the empirical estimator via bootstrapping. Two methods are proposed to fit a distribution to the data, a computationally intensive one that fits the distribution of empirical entropic risk, and a simpler one that fits the tail of the empirical distribution. The approach is applied to a distributionally robust entropic risk minimization problem with type- ∞ Wasserstein ambiguity set, demonstrating improved calibration of size of the ambiguity set using debiased validation performance.

4 - Smart Auctions for Smart Grids: Learning-enabled Local Energy Market Design

Lesia Mitridati, Technical University of Denmark, Kgs Lyngby, Denmark

Citizen energy communities and local energy markets offer a promising pathway to unlock the flexibility of consumers and prosumers. Yet, designing efficient and scalable consumer-centric energy markets remains challenging. Indeed, flexible consumers and prosumers often have heterogeneous preferences that go beyond purely financial considerations, including when and where energy is generated, how it is sourced, and even who they are trading with. Capturing these complex preferences over multiple energy attributes without placing unrealistic computational burdens on participants is key to building practical local energy markets.

This talk presents recent advances, fusing online learning and mechanism design to address these challenges. First, we introduce a novel framework that integrates hierarchical optimization with Thompson sampling to dynamically set optimal prices while simultaneously learning prosumer behaviors through observed responses. We then show how online learning techniques can support the design of scalable iterative combinatorial auctions that match complex supply and demand preferences over energy bundles.

Through realistic numerical studies across simulated energy communities, we demonstrate that these learning-enabled market mechanisms can quickly converge to efficient and fair allocations, while preserving prosumers' autonomy and minimizing communication overhead. Together, these results pave the way toward adaptive consumer-centric local markets for decentralized energy systems.

5 - Structured Learning for TIME Series Aggregation in Net-Zero Power Systems: from Simplex Theory to Scalable Optimization

Sonja Wogrin, Graz University of Technology, Graz, Austria, Yannick Werner, Luca Santosuosso, Beltran Castro Gomez

The transition to net-zero energy systems involves integrating renewables, flexibility technologies, and sector coupling - dramatically increasing the complexity of long-term energy system optimization models that depend on time series input data. The ERC-funded project *NetZero-Opt* proposes a novel time series aggregation (TSA) method that utilizes Dantzig's simplex algorithm for the identification of active constraint sets and provides performance guarantees.

In this presentation, we showcase how machine learning, grounded in simplex foundations, is leveraged within the *NetZero-Opt* framework to accelerate computational performance. We highlight two key applications: (1) using a classifier to predict active time-linking (e.g., ramping or storage) constraints in power system models. This prediction enables us to disaggregate large-scale linear programs and solve parallel submodels, significantly reducing computation times without compromising accuracy. (2) TSA is exact when you aggregate data points from the same set of active constraints, but it incurs an error when merging points from multiple active constraint sets. We use learning to identify optimal mergers to minimize the output error to leverage performance gains.

Together, these learning-based methods with theoretical underpinnings for TSA illustrate a new frontier in scalable optimization for energy system modeling.

MB31

Building B Level 2 B205

Operations Research and Game-Theoretic Approaches for Sustainable Natural Resource and Energy Systems – Part 2

Invited Session

ENRE: Natural Resources

Chair: Sabah Bushaj, SUNY Plattsburgh, Plattsburgh, NY, United States

Co-Chair: Hyunwoo Lee, Virginia Tech, Blacksburg, VA, 24060, United States

1 - Conservation Incentive Design Under Uncertainty: a Stochastic Differential Game Approach with Asymmetric Information

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

Water rights trading (WRT) is a market-based mechanism that promotes conservation in the management of water resources, particularly in response to water scarcity. In a voluntary WRT setting, water users, whose water rights were historically allocated and remain fixed, can temporarily lease or permanently sell them to other users or conservation organizations. This study designs conservation incentives using stochastic differential games with asymmetric information. The problem is set within a system involving three main networks of players: Non-Governmental Organizations (NGOs), groundwater users, and surface water users. The NGO aims to design incentive mechanisms that encourage water users to sell or lease their water rights in response to its offer. Following the NGO's offer, water users have three options: trade their water rights with the NGO, trade with other water users, or choose not to engage in trading. However, the NGO does not have perfect information about the individual water users, such as their initial willingness to participate in the offer or their value of water, while the users have more complete knowledge about their conditions. Additionally, some parameters, such as the actual water demand of users during each time interval (which may fluctuate due to factors like weather), remain uncertain for both the NGO and the users at the beginning of each time interval. Analyzing how these parameters affect the players' behavior is the foundation for the next step in the sensitivity analysis. Finally, The Red River basin serves as the case study.

2 - An Infinitesimal-Interval Approach to Markovian Coherent Risk Measures in Continuous Time

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

We study the formation of the risk function which, in continuous time, assesses the risk borne by a decision maker who has coherent-risk attitudes and faces a finite-state Markov state process. In our limiting process involving ever finer partitions of time, risks are characterized by semi-derivatives that were introduced by Dentcheva and Ruszczyński (2018) to measure coherent risks' changes in response to random state oscillations. Interestingly, convexity of the risk-representing dual sets would play a pivotal role in our pursuit for convergence. The function achieved in the limit is validated by its satisfaction of an ordinary differential equation (ODE) that bears one unique solution. Our infinitesimal-interval approach has the added benefits of relying on only model primitives and providing a natural discrete-time approximation scheme. It can be applied to some cases involving continuous states as well, when transition probabilities are substituted by cumulant generating functions (CGFs). Both possess the "near proportionality to a small time increment" property.

3 - Spatial Decision-Making Under Markovian Dispersal: a Mixed-Integer Programming Approach

Chao Wang, University of Maryland, Cambridge, MD, United States, Jorge Sefair

We develop a spatial optimization framework for ecological corridor design under uncertain animal movement using a Spatial Absorbing Markov Chain (SAMC) model. This formulation captures probabilistic dispersal across fragmented landscapes, where the goal is to select a subset of spatial nodes that forms a movement corridor maximizing the likelihood of successful absorption into habitat patches, subject to budgetary, biological, and geometrical constraints. The objective function is non-additive and defined over transition-based absorption probabilities, leading to a substantially more complex solution structure than conventional deterministic or additive utility models. While the formulation does not fall under classical stochastic programming, it embeds ecological uncertainty via a Markovian process within a nonlinear mixed-integer programming framework, capturing spatial variability without relying on scenario-based enumeration. To address the model's nonconvexity and combinatorial scale, we design a tailored solution method that incorporates problem reformulations, warm-start initialization, valid inequalities, bound tightening heuristics, and variable reduction techniques. We demonstrate the model and algorithm on both synthetic testbeds and a real-world application involving Florida panther habitat connectivity. Results show our approach significantly outperforms commercial solvers and scales effectively to large, heterogeneous landscapes. This work offers a novel bridge between stochastic ecological modeling and large-scale spatial decision-making.

4 - Best-Response Dynamics Incorporated Zero-Regret Algorithm for Large-Scale Integer Programming Games with Application to Aquatic Invasive Species Prevention

Hyunwoo Lee, Virginia Tech, Blacksburg, VA, United States, Robert Hildebrand, Esra Toy, selina cai

This paper presents a scalable algorithm for computing the best pure Nash equilibrium (PNE) in large-scale integer programming games (IPGs). While recent advances in IPG algorithms are extensive, existing methods are limited to a small number of players, typically $n=2,3$. Motivated by a county-level aquatic invasive species (AIS) prevention problem involving 84 players -- we develop efficient and scalable algorithms that significantly extend the applicability of IPGs. Specifically, we propose the best-response dynamics incorporated zero-regret (BZR) algorithm, which leverages best-response dynamics (BRD) for rapid PNE identification and integrates BRD as a primal heuristic within the zero-regret (ZR) framework. This approach dramatically improves the scalability of IPG algorithms, allowing us to solve IPGs with up to 30 players in random datasets and the 84-player AIS prevention problem with Minnesota data. To model the AIS prevention problem, we introduce the Edge-weighted Budgeted Maximum Coverage (EBMC) game, a new class of IPG that has not been previously studied. We establish theoretical conditions for the existence of a PNE under both selfish and locally altruistic utility functions. Experimental results in EBMC games and knapsack problem games demonstrate that BZR significantly enhances ZR in both finding a PNE and identifying the best PNE, particularly in games with many players.

MB32

Building B Level 2 B206

Navigating Sustainable Operations in Energy and Transportation Sectors

Invited Session

ENRE: Environment and Sustainability

Chair: Jiarui Wei, University of Texas at Dallas, Richardson, United States

Co-Chair: Lingling Shi, Macmaster, Hamilton, ON, Canada

1 - Dismissal of Demand Dependence Disappoints and Deceives: the Case of Microgrid Generation and Storage Investments

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

Microgrids (MGs) emerge as vital components of energy transition, offering local and low-cost solutions, e.g., for meeting the increasing demand of data centers and hence relieving the pressure on the grids. MGs, comprising of local generators, storage equipment and consumers, trade electricity with the grid through contracts specifying grid's purchase price. Determining optimal generation and storage capacities for MGs presents significant challenges due to uncertainties in the demand and market price, and specifically their dependence. We incorporate uncertainties and dependences when evaluating and maximizing MG profits. Moreover, we present closed-form solutions for optimal generation and capacity levels. To formulate dependence, we rely on a positively associated representation of demand and price pair. We include empirical analysis to showcase this dependence and a realistic case study along with a sensitivity analysis. We identify weak conditions for generation and storage optimization to bypass traditional dynamic programming approaches. We obtain two major and novel results: ignorance of demand dependence (of price) artificially inflates profits and yields lower optimal capacities. We identify the type of microgrid-grid contract and positively associated demand-price pairs as drivers of our results by studying an alternative contract and negatively associated pairs. We package our key insights as takeaways and customize them for the MG stakeholders. MG operators should capture demand dependence to avoid financial disappointments. In contrast, these operators might ignore demand dependence when negotiating with financial institutions to secure favorable financing terms.

2 - Remanufacture OR Salvage? Reverse Logistics Optimization for Wind Turbines with Imperfect Sorting

Siqiang Guo, University of Missouri-St. Louis, St. Louis, MO, United States, Haitao Li

This study examines reverse logistics decisions for retired wind turbines under imperfect on-site sorting. We model the classification errors using a confusion matrix and develop a cost-minimization framework that captures key trade-offs among remanufacturing profit, salvage value, transportation cost, and sorting accuracy. Closed-form optimality conditions are derived for different sorting scenarios, revealing when remanufacturing is economically justified. Numerical experiments validate the analysis and show how system performance changes with turbine condition and sorting quality. The study addresses a key gap in the literature by jointly considering imperfect sorting and high transportation costs in wind turbine recovery.

3 - Integrating EV Charging and Discharging into Power Grid Through Bilateral Negotiation

Linqiu Li, University of Science and Technology of China, Hefei, China, People's Republic of, Yun Fong Lim, Kai Pan

To deal with demand uncertainty on a power grid, a power plant with limited ramping capability can collaborate with an electric vehicle (EV) company. With proper charging and discharging prices, the EV company voluntarily withdraws electricity from or returns electricity to the power grid in suitable phases. We model the two parties' interactions as a bargaining game on the prices, followed by the EV company's charging and discharging problem and the power plant's electricity generation problem. To solve this bargaining game, we propose a novel "Guess and Verify" approach. Specifically, we first find an optimal solution within a restricted price set in which the two parties' total cost is minimized, and then verify its global optimality. Under an equilibrium contract, we find that the power plant can reduce its expected cost from the collaboration. This is because the EV company fully charges in a low electricity demand phase, reducing the power plant's curtailment cost, and fully discharges to the power grid in a high electricity demand phase, lowering the power plant's electricity generation cost. Based on real data, our numerical experiments suggest that the EV company's charging and discharging can substantially harmonize the power flow within the grid and save significant cost, especially when the electricity demand gap across different phases increases or the power plant's ramping capability decreases. Surprisingly, the EV company's percentage cost saving can exceed 100%, implying that it can make a profit from the collaboration. For the power plant, the percentage cost saving is 2-7%.

4 - Strategic Deployment of Electric Vehicle Charging Infrastructure in South Korea

Na Rea Cho, Wichita State University, Wichita, KS, United States, JINAH YANG, Zahra Ziaei

While electric vehicles (EVs) have garnered worldwide attention, the lack of adequate charging infrastructure remains a significant hurdle to widespread adoption. To address this challenge and support the South Korean government's goal of accelerating EV adoption rates, this study develops a data-driven approach to optimize the placement and capacity of EV chargers in a planned city in South Korea. This study first forecasts EV charging demand based on 7 years of annual demographic attributes and historical charging data. Using these demand projections, we then determine the optimal locations and capacities of charging stations while accounting for government regulations and investment considerations. Our findings identify key drivers influencing infrastructure investment and provide insights into the broader implications of EV charging deployment. Additionally, a case study using data from a Korean city validates the effectiveness of our approach and demonstrates its practical applicability in guiding sustainable urban mobility planning.

MB33

Building B Level 2 B207

Illicit Networks: Modelling, Analysis, and Disruption

Invited Session

Health Applications Society

Co-Chair: Monica Gentili, University of Louisville, Louisville, KY, United States

Co-Chair: Hasini Balasuriya, University of Louisville, Louisville, KY, United States

1 - Uncovering Illicit Supply Chains with Network Optimization: An Illustration Using Cocaine Flow in the United States

Courtney Paulson, University of New Hampshire, Durham, NH, United States, Margret Bjarnadottir, Eunseong Jang, Greg Midgett

Despite increasing crackdowns on illegal drugs, the illegal drug trade in the U.S. alone accounted for approximately \$500 billion in 2024. Combating illicit activities presents a unique challenge for law enforcement officials across the globe, primarily because by their nature, illicit networks are not known or directly observable. Illicit drugs follow hidden supply chains, forcing analysts to rely on secondary data to estimate patterns and movement. This proxy data is naturally limited, inconsistent, and highly variable. The bigger concern for modelers, however, is that these supply chains have no intrinsic ground truth on which to evaluate or train network models. To address these concerns, we develop a novel data-driven optimization approach to combine known data (e.g., law enforcement drug seizure data) with underlying estimation data (e.g., illegal drug fatalities) to maximize learning even with noisy and inconsistent illicit network data. We demonstrate our approach using real-world data on U.S. cocaine seizures and undercover purchases, which we combine with data gathered on overdose deaths. This allows us to derive network structures that are consistent with our data sources, validated using anecdotal data. We further show that we can augment the consistent mathematical modeling of these networks by varying optimization parameters, which create multiple networks based on comparative underlying structures. In the absence of a ground truth, these differing base models offer unique insights for law enforcement agents and governmental policymakers looking to develop forward-thinking strategies to address increasingly complex illicit networks.

2 - Agent-Based Modeling of Opioid use and Harm Reduction Strategies in Community Settings

Melika Jahan Beikloo, Clemson University, Clemson, SC, United States, Hyojung Kang, Mary Beth Kurz, Bryan Lee Miller, Thomas Sharkey, Emily Suiter

The opioid crisis remains a serious public health and criminal justice challenge in the United States, with synthetic opioids like fentanyl contributing to a sharp rise in overdose deaths and the situation further worsened by cuts to addiction treatment funding at state and federal levels. Addressing this complex issue requires tools that can capture the dynamic interactions between individuals, drug markets, and intervention efforts at the community level.

In this work, we present an evolving agent-based modeling (ABM) framework that simulates opioid use progression, drug distribution, and potential harm reduction strategies. Our model includes multiple agent types—such as non-users, opioid users, law enforcement, healthcare systems, and drug dealers—and allows for behavioral transitions, movement, and interactions, such as drug transactions or peer influences that affect individual behavior. We take an iterative approach to model development, continuously refining structure, assumptions, and complexity.

By modeling these dynamics, we aim to support ongoing efforts to better understand opioid-related outcomes and inform the design of effective, data-driven interventions.

3 - Unifying Fragmented Human Trafficking Data: Structural Insights from Ctdc and Unodc Using Ergms

Meng-Hao Li, George Mason University, Arlington, VA, United States, Naoru Koizumi

Both the Counter-Trafficking Data Collaborative (CTDC) and the United Nations Office on Drugs and Crime (UNODC) provide critical data on human trafficking. However, each reflects distinct reporting biases: CTDC captures cases involving victims who receive assistance from partner organizations, while UNODC relies on law enforcement and judicial statistics. These different reporting methods can lead to incomplete and biased views of human trafficking structures due to underreporting or selective detection.

We address this challenge by employing Exponential Random Graph Models (ERGMs) to compare country-level trafficking networks derived from each dataset. By incorporating structural dependencies and node-level covariates, ERGMs allow us to examine whether the two trafficking networks are shaped by similar formation processes. While this comparison highlights structural divergence between datasets, it does not resolve inconsistencies in how trafficking flows are represented.

To bridge this gap, we use a pooled ERGM, a form of multi-network ERGM that jointly models the CTDC and UNODC networks. This approach constrains both datasets to share the same structural parameters while allowing each to differ in baseline density. In doing so, it harmonizes heterogeneous sources and recovers consistent cross-dataset edge patterns while accounting for both structural and reporting heterogeneity. This strategy allows us to treat CTDC and UNODC not as competing accounts, but as complementary views of the same underlying trafficking structure.

4 - Organ Trafficking and Illicit Financial Flows: a Global Study of Kidney Price Spreads

Neelam Shukla, George Mason University, Arlington, VA, United States, Olzhas Zhorayev, Zifu Wang, Melis Ozdemir, Momo Toyoda-Tun, Meng-Hao Li, Naoru Koizumi

The illegal organ trade remains a significant global issue driven by severe shortages in legitimate organ supplies. In 2017, the Global Financial Integrity (GFI) estimated that kidney buyers paid between \$50,000 and \$120,000 per organ, while sellers received less than 10% of this amount. The GFI report primarily relies on aggregated estimates from official sources and lacks detailed transaction-level data, making it difficult to understand the specific factors behind these large price differences. This study aims to elucidate the factors influencing the price differential and construct a novel, transaction-level dataset of kidney trafficking. We systematically identified 978 kidney trafficking-related news articles (2000-2022) using LexisNexis and manually extracted illicit financial data (buyer payments, seller receipts, broker and surgeon fees, transaction and surgery locations). We also explored the use of Llama3.3-70B to extract prices, although the prediction remained poor due to the complex nature of transactions. The final dataset includes 96 transactions involving buyers from 46 countries and 124 involving sellers in 76 countries. Empirical analysis revealed that buyers paid an average of \$76,210 per organ, up to \$300,000, while sellers received an average of \$7,290, with some receiving as little as \$45. This 90% gap aligns with GFI's estimates. We further explored how national factors (e.g., income levels, income inequality, enforcement intensity, rule of law, corruption, social capital, and healthcare accessibility) influence the price spread and found some country-level variation. Overall, this study provides an evidence-based foundation for more targeted international policy responses aimed at reducing exploitation in illicit organ transactions.

MB34

Building B Level 2 B208

Optimization for Statistics and Learning

Invited Session

OPT: Machine Learning

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

1 - A Variational Inequality Approach to Nonlinear Tomographic Reconstruction

Kabir Verchand, University of Southern California, Los Angeles, CA, United States, Ashwin Pananjady

We consider the problem of signal reconstruction for computed tomography (CT) under a nonlinear forward model that accounts for exponential signal attenuation, a polychromatic X-ray source, general measurement noise (e.g. Poisson shot noise), and observations acquired over multiple wavelength windows. We develop a simple iterative algorithm for single-material reconstruction, which we call EXACT (EXtragradiant Algorithm for Computed Tomography), based on formulating our estimate as the fixed point of a monotone variational inequality. We prove guarantees on the statistical and computational performance of EXACT under practical assumptions on the measurement process. We also consider a recently introduced variant of this model with Gaussian measurements, and present sample and iteration complexity bounds for EXACT that improve upon those of existing algorithms. We apply our EXACT algorithm to a CT phantom image recovery task and show that it often requires fewer X-ray projection exposures, lower source intensity, and less computation time to achieve similar reconstruction quality to existing methods.

2 - Rectangularity and Duality of Distributionally Robust Markov Decision Processes

Yan Li, Texas A&M University, College Station, TX, United States, Alexander Shapiro

The main goal of this paper is to discuss several approaches to formulation of distributionally robust counterparts of Markov Decision Processes, where the transition kernels are not specified exactly but rather are assumed to be elements of the corresponding ambiguity sets. The intent is to clarify some connections between the game and static formulations of distributionally robust MDPs, and delineate the role of rectangularity associated with ambiguity sets in determining these connections.

3 - E-Minimax Solutions of Statistical Decision Problems via the Hedge Algorithm

Lezhi Tan, Management Science and Engineering Department, Stanford University, Stanford, CA, United States, Andrés Aradillas Fernández, Jose Blanchet, Jose Luis Montiel Olea, Chen Qiu, Jörg Stoye

We present an algorithm for obtaining ϵ -minimax solutions of statistical decision problems. We are interested in problems where i) the statistician is allowed to choose randomly among I decision rules, and ii) the statistical model may have a parameter space with infinitely many elements. The minimax solution of these problems admits a convex programming representation over the $(I - 1)$ -simplex, and the algorithm suggested herein to obtain an ϵ -approximation of the minimax solution is a version of mirror subgradient descent, initialized with uniform weights and stopped after a finite number of iterations. The resulting iterative procedure is known in the computer science literature as the Hedge algorithm (a particular case of the Multiplicative Weights update method) and it is used in algorithmic game theory as a practical tool to find approximate solutions of two-person zero-sum games. We apply the suggested algorithm to different minimax problems in the econometrics literature. An empirical application to the problem of optimally selecting sites to maximize the external validity of an experimental policy evaluation illustrates the usefulness of the suggested procedure.

4 - Necessary Optimality Conditions in Multiple Hypotheses Testing Under Exact Family-Wise Error Rate Control -- a Path to Fast Computing

Prasanjit Dubey, ISyE Georgia Tech, Atlanta, GA, United States, Xiaoming Huo

In multiple hypotheses testing problem, the computation of the most powerful test is numerically challenging. In the state-of-the-art, the searching for the most powerful test can be done via solving a constrained optimization problem. The dual problem has been derived and strong duality has been established in a very general sense. However, solving the dual problem is still an open question and could result in an exhaustive search, which is not numerically efficient. In this paper, under the exact family-wise error rate control, we proved some necessary conditions for the optimizer of the dual problem. These necessary conditions can lead to coordinate-wise fast algorithms, which may result in numerically efficient algorithms to solve the dual problem, which consequently solve the primal problem for identifying the most powerful test. We prove the linear convergence of our algorithm (that is, the computational complexity of our proposed algorithm is proportional to the logarithmic of the reciprocal of the target error). To the best of our knowledge, this is the first time such a fast algorithm is proposed for finding a most powerful test with family-wise error rate control.

5 - Statistical Estimation via Monotone Variational Inequality for Generalized Linear Models and Generalized Additive Models

Jonghyeok Lee, Georgia Institute of Technology, Atlanta, GA, United States

We propose a statistical estimation framework for generalized linear models (GLMs) based on variational inequalities (VI). Unlike the classical MLE, the VI estimator only requires the link function to satisfy a weak condition (strong Minty's condition), allowing broader applicability even for non-monotone link functions. We establish non-asymptotic estimation error bounds that match or improve upon those of MLE, while relying on weaker assumptions. Furthermore, we identify principles for designing link functions that yield sharper statistical rates and improved algorithmic efficiency. Building on these insights, we develop gradient-based and stochastic iterative methods with provable convergence guarantees. Our results demonstrate that VI estimation provides a flexible alternative to MLE, with advantages in both theoretical guarantees and practical performance across GLMs.

MB35

Building B Level 2 B209

Facility Location Modeling for Critical Infrastructure

Invited Session

Location Analysis

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

1 - Hub Location Problem with a Mixed Green Fleet

Okan Dukkanci, European University Viadrina, Frankfurt (Oder), Germany, James Campbell, Achim Koberstein

This study presents a new hub location problem with a mixed fleet of vehicles, including diesel-based vehicles, electric vehicles, and unmanned aerial vehicles. Each vehicle type has a different cost structure, payload capacity, and traveling range. The objective function minimizes the sum of the recharging (or refuelling) cost, driver cost, fixed vehicle cost, and waiting cost for vehicles and drivers at recharging stations. While the transportation cost (including the recharging cost, driver cost, and vehicle cost) depends on the traveled distance, the recharging fee and waiting cost depend on the number of visits to recharging stations. We develop a mixed-integer linear programming formulation with preprocessing rules and also propose a relax-and-fix heuristic approach. The computational experiments are conducted over well-known CAB and TR data sets, and also a new German data set. The computational results evaluate the performance of the relax-and-fix heuristics and analyze the impact of using different mixed fleets, complete hub networks, different cost structures, and a hypothetical drone with a larger payload capacity and a longer flying range. Results show how the optimal fleet mix, in particular the use of electric vehicles and drones, and the optimal network vary around the world, as these depend on the geographic scope of the services, the distribution of nodes and arc lengths, and the relative costs for fuel (electricity or diesel) and labor.

2 - Citywide Orchestration of Electrical Vehicle Battery Swapping

Zhiyuan Wang, Tsinghua University, Beijing, China, People's Republic of, Xian Guo, Wei Qi, Lun Ran

Electrical vehicle (EV) battery swapping has been developing rapidly in recent years. Despite the significant increase in the number of stored batteries within swapping service systems, mismatches between swapping demand and available batteries remain common and still unresolved. Coordinating the uneven spatial demand and battery inventory, which referred to as spatial coordination, is challenging. One potential solution to the spatial coordination issue arises from the rapid advancement of autonomous driving technology in electric vehicles: coordinated swapping, where EVs perform battery swaps at designated stations based on realized demand, with scheduling managed by a centralized service platform. However, it remains unclear how to design an infrastructure network for coordinated swapping or whether such an approach would lead to cost or battery savings. This paper investigates this emerging urban infrastructure network. Specifically, we model the operations of battery swapping stations using stochastic inventory theory and propose a two-stage distributionally robust optimization model for the joint location and capacity planning of a citywide battery swapping station network. We also propose two alternative strategies in battery swapping designed for specific scenarios, reformulating them as mixed-integer second-order cone programmings. In a numerical study, we validate our algorithm's efficacy using classic transportation data and compare the performance of the three strategies with real transportation data from Beijing. Our findings indicate that the coordinated swapping outperforms the other two strategies in terms of spatial coordination of swapping demand, thanks to its flexibility and economies of scale as the battery swapping network expands.

3 - Optimizing Public Access Automated External Defibrillators (Aed) Deployment Using Smartphone Dispatch Apps

Yifan Wu, Texas A&M University, College Station, TX, United States, Joseph Geunes, Xiaofeng Nie

Motivated by emerging urban applications in commercial, public-sector, and humanitarian logistics, we revisit continuous p-hub location problems under the L1 metric, where several facilities are to be located in a continuous space such that the expected minimum Manhattan travel distance from a random service provider to a random receiver through exactly one hub facility is minimized. Firstly, some closed-form results are derived for one-dimensional cases and two-dimensional cases up to two hubs. Secondly, we propose a simulation-based approximation method for problems with more than two hubs in two-dimensional cases. Finally, we apply our method to a case study on optimizing public access defibrillator locations using smartphone dispatch Apps.

4 - From Charging Costs to Infrastructure: a Cost-Informed Facility Location Model for EV Charging Under Operational Uncertainty

Ehsan Mahyari, The University of Scranton, Scranton, PA, United States

This research presents a novel integration of stochastic operational decision-making into strategic infrastructure planning for electric vehicle (EV) charging systems. Building on our approximate dynamic programming (ADP) framework for managing heterogeneous EV fleets at multi-connector depots, we propose a cost-informed facility location model that leverages operational insights to determine (i) the number of charging depots, (ii) the configuration of chargers at each site, and (iii) their optimal geographic placement. Unlike traditional facility location models that rely on static demand and deterministic costs, our approach embeds ADP-based cost evaluations within a multi-scenario optimization framework. For each candidate site and fleet service region, we simulate the ADP-managed operations under stochastic EV arrival patterns, variable electricity pricing, and charger availability. These simulations yield robust cost-to-serve estimates that inform the facility location decisions, enabling planners to minimize long-run charging costs while meeting service-level guarantees and grid constraints. The resulting model captures the interplay between operational policy and infrastructure design, allowing Charging-as-a-Service (CaaS) providers and public agencies to identify infrastructure configurations that are both resilient and cost-effective under real-world uncertainty. Numerical experiments based on urban and regional fleet scenarios demonstrate that the proposed approach significantly reduces total system cost and unmet demand compared to decoupled or static planning methods.

5 - Disaster Sheltering Network Design Prioritizing Welfare Enhancement and Public Health Protection

Mohamed Salama, Purdue University, West Lafayette, IN, United States, Min Kyung Lee, Yuehwen Yih

Although public shelters are crucial for safeguarding disaster survivors, many individuals risk their safety by refusing to evacuate. This is often due to the inferior conditions typically found in public shelters, such as crowding, continuous lighting, lack of privacy, and safety apprehensions, combined with the absence of alternative sheltering options. Additionally, the risk of infectious disease transmission, such as tuberculosis and COVID-19, further discourages seeking refuge in these facilities. To

address these issues, this study introduces a four-echelon network design aimed at providing small-capacity shelters that enhance welfare and public health. A two-stage stochastic mixed-integer linear programming model is developed to optimize the selection of opening nodes across the four echelons under various disaster scenarios, in addition to determining the associated logistical operations. This model considers the tradeoff between maximizing welfare-enhanced sheltering occupancy and minimizing related costs. To efficiently solve the model and offer practical benefits, a novel problem-specific scenario-based cutting-plane algorithm is developed.

MB36

Building B Level 2 B210

SpORts II

Invited Session

spORts

Chair: Keith Willoughby, University of Saskatchewan, Saskatoon, SK, Canada

1 - Ranking Sports Teams, Teaching Analytics, and Reaching the Public

Laura Albert, University of Wisconsin-Madison, Madison, WI, United States

In this talk, we will discuss how to rank sports teams using a Markov chain-based methodology, which has been used on Badger Bracketology to rank college football, college basketball, and NFL teams. We will present the ranking methodology, with a special focus on its application to the College Football Playoff. Additionally, this talk will highlight how sports analytics can engage students and connect with the public.

2 - Performance Under Pressure: Soccer Penalties and Basketball Free Throws

Nils Rudi, Yale School of Management, New Haven, CT, United States, Andrea Wang

This work considers generative-model approaches to high-pressure scoring events in soccer penalty shootouts and basketball free throws. By specifying a probabilistic process for how each shot arises—capturing latent abilities, sequential decision steps, and contextual pressure—the models can leverage structural assumptions to achieve meaningful inference even with limited data. Bayesian estimation is employed to propagate uncertainty through to latent skill estimates. We will highlight the conceptual parallels between the two sports and discuss how generative assumptions can support performance analytics and strategic decision-making when observations are limited.

3 - A Swimming Coach's Beautiful Mind? Lineup Strategies and Measuring a Swimmer's Value

Matthew Bailey, Bucknell University, Lewisburg, PA, United States

We present a game-theoretic framework to strategically select optimal swim team lineups for competitive meets. By integrating mathematical modeling and game theory principles, we address the decision-making process involved in creating meet lineups by assigning athletes to their events. We seek to maximize the team's expected points, taking into consideration each swimmer's strengths, weaknesses, and event preferences relative to the opponent's potential lineups. The study further analyzes the marginal impact of each selected swimmer on the overall team performance. Our research contributes to a more informed and strategic approach offering decision support for assembling swimming team lineups, while offering coaches valuable insights into maximizing their team's competitive advantage.

4 - The Source of All Frustration: Exploring Schedule Inequities in Canadian Football

Keith Willoughby, University of Saskatchewan, Saskatoon, SK, Canada

Teams in professional sports like baseball, basketball, and hockey may play multiple games per week. However, football teams typically compete in at most one game during a week. Games in the Canadian Football League (CFL) are generally scheduled on Thursday, Friday, Saturday, and Sunday of each week. This forces CFL franchises to play with more (or less) days rest than their opponents. We have engaged with the CFL in building operations research models to inform schedule creation and can confirm that teams view “short weeks” as pain points in the scheduling process. Exploring two decades worth of league schedules, we find some evidence that teams with more days rest than their opponents experience a competitive advantage.

5 - Does Fatherhood Impact the Performance of Professional Cyclists?

Jeroen Belien, KU Leuven, Brussel, Belgium, Anke Baetens, Filip Van den Bossche

This study examines the impact of fatherhood on the performance of professional road cyclists. Fixed effects panel regression is used to compare cycling performance over a period after having a child with the performance in the same period the year before. The results of this study may help to provide better psychological and athletic support for recent fathers and are potentially relevant to contexts other than professional (cycling) sport.

MB37

Building B Level 2 B211

Decision-Making Under Uncertainty: Robustness, Fairness, Interpretability, and Personalization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Phebe Vayanos, University of Southern California, Los Angeles, CA, United States

Co-Chair: Ryan Edmonds, University of Southern California, Los Angeles, CA, United States

1 - Learning Confounding Robust Optimal Risk Scores from Observational Data

Yingxiao Ye, University of Southern California, Los Angeles, CA, United States, Andres Gomez, Cagil Kocyigit, Sina Aghaei, Phebe Vayanos

We consider the problem of learning, from observational data, a logistic regression model to predict the risk of an adverse outcome under no treatment. This problem arises routinely in public health and the social sciences, e.g., to help prioritize individuals for scarce resources or services. The vast majority of the literature assumes unconfoundedness, requiring that we have access to all individual characteristics that were used historically to decide on treatment assignments. However, this assumption often fails to hold in practice, where treatment decisions are often the result of a combination of observed and unobserved characteristics with e.g., conversations that are not recorded anywhere impacting decisions. In this research, we propose a logistic regression-based risk prediction model that is robust to potential unobserved confounders. We demonstrate the effectiveness of our approach on both synthetic and semi-synthetic data, showing that our approach achieves better performance compared to both baseline and state-of-the-art approaches. We also illustrate the power of our approach on the problem of learning a vulnerability score for prioritizing individuals experiencing homelessness for scarce resources and services based on the real data from the LA County Homeless Management Information System.

2 - Learning Robust Prescriptive Trees Under Unobserved Confounders

Ryan Edmonds, University of Southern California, Los Angeles, CA, United States, Phebe Vayanos

We consider the problem of learning prescriptive trees that are robust to unobserved confounders in historical data. This problem arises frequently in high stakes settings such as public health and housing allocation, where policy decisions made in the past may have depended on factors not present in the historical data (i.e., personal interactions between social workers and individuals receiving treatment). These domains also frequently require that learned policies be interpretable, as opposed to black box methods. We propose a method for learning robust prescriptive trees using inverse propensity weighting. In particular, we formulate the problem as a max-min optimization, with the goal of learning a tree robust to adversarial choices of confounding. We show that under some assumptions on how unobserved confounders can affect the true propensity weights, our optimization can be recast as a single stage mixed integer optimization problem. Our formulation has the ability not only to model domain specific constraints, in accordance with previous work on decision trees, but also to model domain specific assumptions on confounders.

3 - A Maximum Likelihood Approach to Randomized Trees

Zhiyuan Sun, University of California, Los Angeles, Los Angeles, CA, United States, Velibor Misic

Randomized trees, unlike traditional deterministic trees, offer a novel framework for predictive modeling by splitting data points with a probability that is determined by their feature vectors and specific parameters at each node. This paper introduces a smooth non-convex Maximum Likelihood Estimation (MLE) problem for learning the parameters of randomized trees from data. We establish that optimizing over randomized trees is theoretically equivalent to optimizing over deterministic trees. We further develop an algorithm that leverages the Expectation-Maximization (EM) method to efficiently optimize the parameters of randomized trees. We theoretically prove the consistency of our randomized tree MLE problem and develop generalization guarantees for the log likelihood of randomized trees by analyzing the Rademacher complexity. Our numerical experiments demonstrate the effectiveness of our method in improving predictive performance over classical top-down induction methods such as CART over a variety of binary and multiclass classification datasets.

4 - Equitable Auction Design: with and Without Distributions

Napat Rujeerapaiboon, National University of Singapore, Singapore, Singapore, Ruiqin Wang, Cagil Kocyigit

We study a mechanism design problem where a seller aims to allocate a good to multiple bidders, each with a private value. The seller supports or favors a specific group, referred to as the minority group. Specifically, the seller requires that allocations to the minority group are at least a predetermined fraction (equity level) of those made to the rest of the bidders. Such constraints arise in various settings, including government procurement and corporate supply chain policies that prioritize small businesses, environmentally responsible suppliers, or enterprises owned by historically disadvantaged individuals. We analyze two variants of this problem: stochastic mechanism design, which assumes bidders' values follow a known distribution and seeks to maximize expected revenue, and regret-based mechanism design, which makes no distributional assumptions and aims to minimize the worst-case regret. We characterize a closed-form optimal stochastic mechanism and propose a closed-form regret-based mechanism, and establish that the ex-post regret under the latter is at most a constant multiple (dependent on the equity level) of the optimal worst-case regret. We further quantify that this approximation constant is at most 1.31 across different equity levels. Both mechanisms can be interpreted as set-asides, a common policy tool that reserves a fraction of goods for minority groups. Furthermore, numerical results demonstrate that the stochastic mechanism performs well when the bidders' value distribution is accurately estimated, while the regret-based mechanism exhibits greater robustness under estimation errors.

5 - Prefactual Recommendation by Bayesian Preference Elicitation

Viet Anh Nguyen, Chinese University of Hong Kong, Hong Kong, Hong Kong

Algorithmic prediction models increasingly influence high-stakes decision-making, highlighting the need to empower individuals to overturn unfavorable predictions. Prefactual recommendations, or algorithmic recourses, provide actionable guidance to help applicants achieve favorable algorithmic predictions with a minimal effort in modifying their features. However, most existing methods assume that all applicants share the same cost function and do not account for subjective differences in the cost function between applicants. To address these shortcomings, we propose the Prefactual Recommendation by Bayesian Preference Elicitation (PROBE) framework, which personalizes prefactual recommendations by learning an applicant's cost preferences. PROBE consists of two principal components: Bayesian preference elicitation and graph-based risk-averse recommendation generation. In the preference elicitation stage, PROBE learns a Mahalanobis approximation of the applicant's cost function through an interactive question-answering scheme. In each round, the most informative questions are selected by maximizing the mutual information between the applicant's response and the prior belief of the cost parameters. Upon observing the applicant's responses, the posterior belief is updated using a variational Bayesian approach, which balances the prior-posterior distortion and the alignment of the posterior with the observed responses. To generate prefactual recommendations, PROBE constructs a graph based on available data to capture actionability constraints and assigns a stochastic edge cost by leveraging the terminal Mahalanobis posterior belief. PROBE then searches for an optimal mean-variance path from the applicant's current feature values to a node with a favorable prediction. This is a joint work with Lemn Kong, Marshal Sinaga and Anthony So.

MB38

Building B Level 2 B212

Advances in Network Interdiction: Theories and Applications

Invited Session

OPT: Network Optimization

Chair: Thanakrit Piyachayawat, Texas tech university, Lubbock, TX, United States

1 - Robust Network Enhancement

Thanakrit Piyachayawat, Texas Tech University, Lubbock, TX, United States, Ningji Wei

Network systems such as critical supply chains, power grids, and water distribution systems are constantly facing challenges due to natural disasters and adversarial disruptions. Enhancements that attempt to improve the associated operational efficiency become critical and challenging due to uncertain factors such as economic fluctuations and investment inefficiency. In this work, we propose a general mathematical model for these network structures while mitigating uncertainties from multiple sources. Through this approach, we obtain effective and robust network enhancement plans that account for the performance range of the network structures, defined by their worst-and best-case scenarios. We develop exact solution methods using cut-generation techniques. To improve the scalability and efficiency, we develop an approximation method with performance guarantee. These findings offer promising results for tackling real-world network enhancement problems under uncertain factors.

2 - Protecting Humanitarian Logistics: An Interdiction Framework for Problems with Shared Resource Interdependencies

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

Modern supply and infrastructure systems are increasingly vulnerable to disruption, both accidental and adversarial, especially when protection and service operations must compete for limited resources. In this work, we study a class of **multi-stage interdiction games** that model such interdependent and adversarial settings through a finite, sequential, and alternating decision process between a *protector*, *attacker*, and *defender*. Specifically, we consider a **facility location allocation problem** in which a shared pool of resources, such as service crews or vehicles, must be allocated across two competing objectives: (1) fortifying regions against potential interdiction, and (3) delivering supplies to customers in the aftermath of an attack.

This inherently creates a **resource trade-off** between stages, where protection decisions made early in the game directly constrain the operational capabilities available for service delivery later. Moreover, interdiction actions taken in the second stage are adaptive to the observed protection policy, compounding the strategic complexity. We formalize this as a three-stage optimization game with interdependencies across non-consecutive stages, where early protection choices can render later service plans infeasible or suboptimal.

To solve these challenging problems, we propose an **exact combinatorial branch-and-bound framework** tailored to capture the structure of these interdependencies. Our approach efficiently explores the space of joint protection and service strategies under adversarial threat. We illustrate the effectiveness of our method through applications to **network-based facility allocation problems**, with relevance to **humanitarian logistics**.

3 - Network Interdiction with General Integer Variables

Anh Truong, University of Florida, Gainesville, FL, United States, Leonardo Lozano, Jorge Sefair

We study a network interdiction problem in which a defender seeks a minimum-cost flow through a capacitated, directed network, while an attacker aims to maximize this cost. The attacker disrupts the network through interdiction actions such as reducing arc capacities, increasing arc costs, or both. The attacker actions are modeled using general integer variables that capture the location and intensity of disruptions. We develop a single-level dual reformulation of the problem, strengthened with three families of valid inequalities, including enhancements tailored for instances with sparse flow and large arc capacities. To improve scalability, we integrate the reformulation with a progressive approximation scheme that reduces network size. Computational experiments on grid networks demonstrate the effectiveness of our approach, expanding the class of tractable network interdiction problems.

4 - An Exact Algorithm for the Bi-Objective Shortest Path Interdiction Problem

Tanner Staggs, University of Florida, Gainesville, FL, United States, Jorge Sefair

We study a shortest-path interdiction problem in which the adversary's behavior (inner level) is modeled as a bi-objective shortest path problem (B-SPI). The leader (outer level) seeks to increase the cost of selected arcs, subject to a budget constraint. We propose a solution strategy that reformulates the problem as the interdiction of a Pareto-optimal set. This approach is embedded within an exact algorithm that decouples the interdiction model from the bi-objective shortest path problem. Our algorithm converges to the optimal solution while evaluating only a subset of the adversary's feasible paths. Computational experiments on two networks of different scales with randomly generated data demonstrate the method's efficiency. We also present a case study on human trafficking interdiction using real road network data from California, highlighting the approach's practical value even when the leader is only able to disrupt one of the adversary's objectives.

5 - Logistic Planning Under Vehicle Interdiction with No-Recourse

Abdulsamed Kagit, Georgia Institute of Technology, Atlanta, GA, United States, Mathieu Dahan, Oktay Gunluk

We study a logistic planning problem on an acyclic directed network where vehicles are subject to interdiction. In this problem, different types of commodities are supplied from a subset of nodes and demand nodes require specific amounts of these commodities. Each vehicle is dispatched from its origin node and has a fixed capacity. A unit cost is incurred for each unit of unmet demand at a demand node while no cost is incurred for excess supply.

The logistic planner first determines the vehicle routes, specifying through which arcs and by which vehicles the commodities are transported. After the logistics plan is established, the attacker selects an interdiction strategy that disables vehicles on the arcs they traverse. Interdicted vehicles are removed from the network along with the commodities they were carrying. Interdicting a vehicle on an arc incurs a cost and the attacker must remain within a given budget. In the no-recourse setting we consider, the logistic planner cannot revise the plan after interdictions are realized. Only the originally planned vehicle routes that remain intact can be used to deliver commodities. The logistic planner's objective is to minimize the total cost incurred from unmet demand in the worst-case scenario while the attacker's objective is to maximize it. We present a single level reformulation of this problem and develop a branch-cut-and-price algorithm to solve it.

MB39

Building B Level 2 B213

Quantum Computing and Optimization

Invited Session

OPT: Linear and Conic Optimization

Chair: Pouya Sampourmahani, Lehigh University, Bethlehem, PA, United States

Co-Chair: Mohammadhossein Mohammadisiahroudi, Lehigh University, Hatfield, 19440, United States

1 - Quantum Computing Inspired Iterative Refinement for Semidefinite Optimization

Tamás Terlaky, Lehigh University, Bethlehem, PA, PA, United States, Mohammadhossein Mohammadisiahroudi, Pouya Sampourmahani, Zeguan Wu

Iterative Refinement (IR) is a classical computing technique for obtaining highly precise solutions to linear systems of equations, as well as linear optimization problems.

The limited precision of quantum solvers motivated us to develop the first IR scheme for solving semidefinite optimization (SDO) problems and explore two major impacts of the proposed IR scheme. First, we prove that the proposed IR scheme exhibits quadratic convergence of the optimality gap without any assumption on problem characteristics. As an application of our results, we show that using IR with Quantum Interior Point Methods (QIPMs) leads to exponential improvements in the worst-case overall running time of QIPMs, compared to previous best-performing QIPMs. We also discuss how the proposed IR scheme can be used with classical inexact SDO solvers, such as classical inexact IPMs with conjugate gradient methods.

2 - Almost-Exact Quantum Interior Point Method for Linear Optimization

Pouya Sampourmahani, Lehigh University, Bethlehem, PA, United States, Mohammadhossein Mohammadisiahroudi, Zeguan Wu, Jun-Kai You, Tamás Terlaky

The emergence of huge-scale, data-intensive linear optimization (LO) problems in applications such as machine learning has driven the need for more computationally efficient interior point methods (IPMs). While conventional IPMs are polynomial-time algorithms with rapid convergence, their per-iteration cost can be prohibitively high for dense large-scale LO problems. Quantum linear system solvers have shown potential in accelerating the solution of linear systems arising in IPMs.

In this work, we introduce a novel almost-exact quantum IPM, where the Newton system is constructed and solved on a quantum computer, while solution updates occur on a classical machine. Additionally, all matrix-vector products are performed on the quantum hardware. This hybrid quantum-classical framework achieves an optimal worst-case scaling of $O(n^2)$ for fully dense LO problems. To ensure high precision, despite the limited accuracy of quantum operations, we incorporate iterative refinement techniques both within and outside the proposed IPM iterations.

The proposed algorithm has a quantum complexity of $O(n^{1.5} \kappa_A \log(1/\epsilon))$ queries to QRAM and $O(n^2 \log(1/\epsilon))$ classical arithmetic operations. Our method outperforms the worst-case complexity of prior classical and quantum IPMs, offering a significant improvement in scalability and computational efficiency.

3 - Quantum-Inspired Hamiltonian Descent and Real-World Applications

Yuxiang Peng, Purdue University, West Lafayette, IN, United States, Jiaqi Leng, Samuel Kushnir, Jinglei Cheng, Pengyu Liu, Hanrui Wang, Xiaodi Wu

Quantum Hamiltonian Descent (QHD), a recent quantum algorithm, has been theoretically shown to achieve exponential speedups in certain optimization problems and has empirically outperformed state-of-the-art classical solvers on small-scale benchmarks. However, the limited scalability of current quantum hardware poses significant challenges for deploying QHD in real-world applications.

Motivated by the quantum and classical mechanics principles underpinning QHD, we introduce Quantum-Inspired Hamiltonian Descent (QIHD)—a scalable, classical variant that retains the core advantages of QHD while being suitable for large-scale optimization tasks. QIHD embeds optimization objectives into a classical Hamiltonian system derived from the dequantization of QHD's quantum dynamics, enabling more effective exploration of complex landscapes in large-scale and data-intense optimization problems compared to conventional methods.

In this talk, I will present the theoretical foundations of QIHD, its efficient GPU-based implementation, and its empirical performance across a range of benchmarks. I will also highlight real-world applications in various domains, where QIHD demonstrates better efficiency and higher solution quality compared to existing approaches. This work illustrates how quantum algorithms can inspire the development of modern optimization solvers with broader applicability.

4 - QAOA-GPT: Efficient Generation of Adaptive and Regular Quantum Approximate Optimization Algorithm Circuits

Ilya Safro, University of Delaware, Newark, DE, United States

We introduce QAOA-GPT, a generative framework that leverages Generative Pretrained Transformers (GPT) to directly synthesize quantum circuits for solving quadratic unconstrained binary optimization problems, and demonstrate it on the MaxCut problem on graphs. To diversify the training circuits and ensure their quality, we have generated a synthetic dataset using the adaptive QAOA approach, a method that incrementally builds and optimizes problem-specific circuits. The experiments conducted on a curated set of graph instances demonstrate that QAOA-GPT, generates high quality quantum circuits for new problem instances unseen in the training as well as successfully parametrizes QAOA. Our results show that using QAOA-GPT to generate quantum circuits will significantly decrease both the computational overhead of classical QAOA and adaptive approaches that often use gradient evaluation to generate the circuit and the classical optimization of the circuit parameters. Our work shows that generative AI could be a promising avenue to generate compact quantum circuits in a scalable way.

5 - An Exclusive-Sum-of-Products Pipeline for Grover Mixer Qaoa

Rebekah Herrman, University of Tennessee Knoxville, Knoxville, TN, United States, Matthew Brunet, Shilpi Shah, Mostafa Atallah, Anthony Wilkie

The Quantum Approximate Optimization Algorithm (QAOA) is a quantum algorithm used to approximately solve combinatorial optimization problems. Grover mixer QAOA (GM-QAOA) is a variant of the algorithm in which one can constrain QAOA to only search a feasible solution space instead of the space of all bit strings of length n . However, GM-QAOA requires a specialized initial state that depends on the feasible solution state. This initial state can be challenging to create on quantum hardware. In this work, we discuss an exclusive-sum-of-product approach for creating the GM-QAOA initial states and analyze the performance of the approach.

MB40

Building B Level 2 B214

Human AI Interaction in Retailing

Invited Session

Service Science

Chair: Kejia Hu, University of Oxford Said Business School, Oxford, United Kingdom

Co-Chair: Demei Kong, N/A

1 - Optimizing a Human-AI Warehouse: Integrating Product Stowage and Retrieval Using Autonomous Robots

Yun Fong Lim, Singapore Management University, Singapore, Singapore, Yuyu Liu, Marcus Ang, Kai Pan, Yugang Yu

We study an autonomous robotic system, such as a Kiva system, operating within a warehouse with multiple storage locations and multiple work stations over a multi-period horizon. Products arrive according to a predetermined schedule but their demands are uncertain. At the start of each period, we select pods from the warehouse to stow each arriving product at each station. Likewise, at the end of the period, we select pods to retrieve products that are demanded at each station. Our objective is to minimize the expected total travel cost of the pods. We formulate the integrated product stowage and retrieval problem over multiple periods with demand uncertainty as a stochastic optimization problem. To make the problem tractable, we partition the warehouse into several storage classes. We introduce a novel two-level hierarchical framework to solve the problem. In Level 1, we first choose classes to stow or retrieve products in each period. In Level 2, we select pods from the chosen classes to stow or retrieve the products in each period. This hierarchical framework scales effectively to large problem instances while yielding practical managerial insights. Numerical experiments suggest that our approach outperforms existing policies by about 49% on synthetic data and 26% on real data from an electronic equipment manufacturer. We find that when the warehouse has a sufficient number of storage classes, class assignment has a greater impact on performance than pod selection, and over partitioning the classes is unnecessary. However, pod selection becomes more critical when pod capacity is limited.

2 - Early Prediction of Product Success Through Integrating E-Commerce and Social Media Dynamics

Qili Wang, University of Florida, Gainesville, FL, United States, Yi Su, Liangfei Qiu, Hsing Kenneth Cheng

This study aims to predict product success to inform design and operational decisions during the early stages of product development. A key challenge at this stage lies in determining which features to include in a new product. Prior research in consumer behavior suggests that product performance is significantly influenced not only by individual features, but also by interactions among features and between features and the product itself. To model these complex relationships, we propose a graph-based data mining framework that integrates data from multiple word-of-mouth sources using text mining techniques. Our framework constructs a heterogeneous graph to capture product-to-product, product-to-feature, and feature-to-feature relationships. Experimental results show that our approach significantly outperforms existing benchmarks, demonstrating strong predictive capabilities and offering valuable practical implications for product development.

MB41

Building B Level 3 B301

HAS Distinguished Speaker Session 2

Invited Session

Health Applications Society

Chair: Pengyi Shi, Purdue University, West Lafayette, IN, United States

Co-Chair: Lauren Steimle, Georgia Institute of Technology, Atlanta, GA, United States

1 - Newborn Screening for Cystic Fibrosis

Ebru Bish, University of Alabama, Tuscaloosa, AL, United States

In this talk, I will present several projects on newborn screening (NBS) for cystic fibrosis (CF), a life-threatening genetic disorder. With over 1,000 known CF-causing variants, comprehensive genetic testing for CF NBS is both costly and capacity-limited. As a result, the U.S. screening protocol begins with a low-cost biochemical test measuring immunoreactive trypsinogen (IRT), which has limited efficacy. Only newborns with IRT levels above a set threshold proceed to genetic testing. I will introduce novel screening designs that can improve both the accuracy and efficiency of NBS. In addition to the practical implications of these designs, I will share insights into their real-world impact and outline ideas for future research inspired by this work.

2 - Operationalizing AI in Healthcare Delivery

Carri Chan, Columbia Business School, New York, NY, United States

The excitement about the potential of AI to improve healthcare delivery and lower costs continues to grow with ever increasing sophistication of data analytics tools. The healthcare industry has focused less on how to actually operationalize and make decisions based on their predictive models. This talk will discuss models, empirical evidence, and practical obstacles to overcome in order to achieve lower costs, better outcomes, and increased access to care.

MB42

Building B Level 3 B302

Empirical Research in Healthcare Analytics

Invited Session

Health Applications Society

Chair: Jong Myeong Lim, Miami Herbert Business School, Coral Gables, FL, 33146, United States

1 - Proactive Admission Control for Emergency Departments: Integrating Demand and Inpatient Admission Delay Predictions

Jing Dong, Columbia University, New York, NY, United States, Ruicheng Ao, Xiaole (Alyssa) Liu, Martin Copenhaver

Emergency department crowding has severe consequences for patient outcomes and staff well-being. Addressing this challenge requires not only reactive measures but also proactive strategies that anticipate demand and admission delays. Advances in machine learning and predictive analytics now allow for more accurate forecasting of arrival rates, inpatient bed availability, and system bottlenecks, providing an opportunity to design data-driven interventions to enhance patient flow and mitigate crowding. In this work, we study how to integrate predictive demand and admission delay information into ED admission control decisions. By proactively regulating the acceptance of transfer patients based on real-time congestion levels and future system conditions, our approach provides a viable way to reduce delays and enhance overall efficiency in emergency care delivery.

2 - Generative AI in Clinical Documentation: Impact on Healthcare Efficiency and Provider Well-Being

Ragip Gurlek, Emory University, Atlanta, GA, United States, Diwas KC

This study empirically examines the effects of ambient documentation tools that automatically generate clinical notes from patient-provider conversations. Using panel data from a large hospital, we identify positive impacts on provider productivity and well-being. Results show these effects are moderated by contextual factors and provider characteristics.

3 - Place and Health: Built Environment and Healthcare Process Outcomes

Eric Xu, Mississippi State University, Mississippi State, MS, United States, RJ Niewoehner, Kevin Linderman

In recent years, the Center for Medicaid and Medicare Services has implemented a Hospital Readmissions Reduction Program to streamline processes and ensure quality outcomes. However, this reductionist approach to treatments has witnessed variable success in terms of patient outcomes. Therefore, we seek to understand the patient context in greater depth as to explain the variable outcomes in a group of patients with heart disease. This study uses patient data from the state of New York to understand how built environment contextual variables can impact treatment process outcomes.

4 - Operational Drivers of Nurse Turnover: Evidence from Operating Room Scheduling Data

Jaeyoung Kim, University of Dayton, Dayton, OH, United States, Ahmet Colak, Lawrence Fredendall

High turnover among operating room nurses poses critical challenges to care continuity and hospital performance. This study examines how the structure of work schedules influences nurse turnover in high-intensity care settings. Using granular hospital data and a pre-trained language model, we characterize nurses' work experiences in terms of their task and team assignments. We aim to quantify how scheduling patterns relate to nurse turnover, offering insights for improving workforce stability through data-informed management strategies.

MB43

Building B Level 3 B303

Decision Intelligence in Healthcare: Operations, AI, and Optimization

Invited Session

Health Applications Society

Chair: Sidian Lin, Harvard University, Cambridge, MA, United States

Co-Chair: Jacob Jameson, Harvard University, Cambridge, MA, United States

1 - From Prediction to Policy - AI-Enabled Decision-Making in Fertility Treatment

Bogdana Jelic, University of Oxford, Oxford, United Kingdom, Agni Orfanoudaki, Holly Wiberg, Gurpreet Kalra

In recent years there has been rising concerns about fertility in the adult population worldwide. The World Health Organization estimates that 48.5 million to 186 million individuals may be infertile globally. In Vitro Fertilisation (IVF) procedure continues to represent one of the best treatments for parents waiting for a child, but it remains highly challenging in terms of financial cost, and difficulty to navigate through the complex decision making process, which is often emotionally and mentally overwhelming for the patients. Undergoing fertility treatments are known to be psychosocially demanding, at times with devastating consequences, including experience of loss, low self-esteem, and stigma. Approximately one-third of women and one-tenth of men with unresolved infertility get a psychiatric diagnosis, showing that the question of quick and more accessible fertility treatments is of essential importance. In recent years, there has been a rise in machine learning and AI models being leveraged in the IVF procedure, used for estimating various quantities of interest. But can it aid the patients, or can uncertainty in the models and induced incorrect estimates further impact patients' emotional difficulty? Our work aims to answer both questions. By using dynamic decision making processes we set up a framework for evaluating what is the optimal decision path for a patient, in terms of time-to-a-baby and cost sensitivity of a particular couple. At the same time, we discuss and establish uncertainty and error bounds for the misspecification of AI model estimations, and we establish the impact of such errors on patient-relevant measures.

2 - Biomarker-Guided Immunotherapy in Metastatic Melanoma

Jacob Jameson, Harvard University, Cambridge, MA, United States

PURPOSE: The treatment landscape for metastatic melanoma has been transformed by immunotherapy, with two primary first-line options: PD-1 inhibitor monotherapy or combination PD-1/CTLA-4 blockade. While combination therapy (nivolumab+ipilimumab) shows improved response rates compared to PD-1 monotherapy, it substantially increases both toxicity and costs. However, a subset of patients demonstrates intrinsic resistance to PD-1 monotherapy, experiencing early progression and poor outcomes. Here, we quantify the health economic value of implementing this biomarker-guided immunotherapy selection strategy.

METHODS: Building on the validated modified decision tree (MDT) model from Tarantino et al. that uses tumor genomic heterogeneity and ploidy to predict intrinsic resistance to PD-1 monotherapy, we developed a partitioned survival model calibrated to 10-year data from CheckMate-067. The model incorporates three health states (progression-free, progressed, dead) with parametric spline functions fit to overall survival and progression-free survival data to evaluate the impact of biomarker-guided therapy selection.

RESULTS: The MDT model shows that patients predicted to be resistant to PD-1 monotherapy have significantly worse survival on monotherapy, but notably, 43% responded to combination therapy, suggesting substantial clinical benefit from appropriate treatment selection. Our survival modeling demonstrates that biomarker-guided therapy selection can improve quality-adjusted survival by directing resistant patients to combination therapy while sparing responsive patients from unnecessary toxicity. Preliminary economic analyses suggest favorable cost-effectiveness of this precision medicine approach compared to standard unguided treatment selection.

CONCLUSIONS: Implementation of this genomic biomarker-guided approach to immunotherapy selection shows promise for optimizing clinical outcomes in metastatic melanoma by matching patients to the most appropriate first-line therapy.

3 - Prediction of Response to Intra-Articular Injections of Hyaluronic Acid for Knee Osteoarthritis

Eva Lee, The Data and Analytics Innovation Institute, Atlanta, GA, United States

Osteoarthritis (OA) is a degenerative joint disease, most commonly affecting the knee. About half of knee OA patients ultimately undergo surgery, though many procedures may be avoidable. Intra-articular injections of hyaluronic acid (HA) offer a non-invasive, cost-effective alternative that improves knee function, yet treatment efficacy remains debated. Outcomes often depend on patient characteristics such as age, weight, gender, and OA severity, as well as injection site and placement.

This study employs a multi-stage, multi-group machine learning model to identify features that predict patient response to various HA treatments. The algorithm uncovers subgroups of patients most likely to benefit from HA therapy. Baseline results highlight factors such as weight, smoking status, and smoking frequency as key predictors, enabling accurate identification of patients best suited for HA injections. The model achieves over 89% blind prediction accuracy.

These findings support more selective administration of HA, improving therapy success rates while reducing unnecessary procedures. Predicted response information can also be shared with patients to align treatment with their values and preferences. Additionally, decision-support tools derived from the model allow physicians to monitor whether patients are meeting expected response levels and, if not, to consider corrective action.

To our knowledge, this represents the first machine learning approach to predicting responses to HA injections for knee OA. The framework is generalizable and may be extended to guide treatment decisions for other conditions.

MB44

Building B Level 3 B304

AI-Driven Decision Analysis: Towards Human and Skills Optimization in the Digital Era

Invited Session

Decision Analysis Society

Chair: Irad Ben Gal, Tel Aviv University, TEL AVIV, Israel

Co-Chair: Hila Chalutz-Ben Gal, Bar-Ilan University, Ramat Gan, Israel

1 - Social Federated Learning (Sfl):: New Insights and Examples**Irad Ben-Gal, Tel Aviv University, Tel Aviv, Israel**

Federated Learning offers a framework for collaborative model training without sharing raw data, thereby addressing critical privacy and communication costs overcoming challenges in distributed environments. This research introduces a novel data sharing scheme that leverages Social Federated Learning (SFL) to enhance performance in non-industry settings. Examples include: Healthcare, Cyber security, and industrial IoT. Empirical evaluations, using synthetic and real-world datasets, demonstrate that our scheme preserves high levels of model accuracy of centralized baselines, while providing strong guarantees on interpretability and privacy.

2 - Social Media Impact and Smartwatch Monitoring: Prevalence and Early Markers of Ptsd and Anxiety Following Mass Traumatic Events**Erez Shmueli, Tel-Aviv University, Tel-Aviv, Israel, Dan Yamin**

We investigated PTSD, anxiety, and stress in individuals indirectly exposed to the October 7 events in Israel. Our study utilized data from a three-year prospective study of 4,806 smartwatch users who completed daily questionnaires, and a panel study surveying 2,536 individuals twice. Stress symptoms were measured daily, while PCL-5 and GAD-7 questionnaires assessed PTSD and anxiety prevalence. After October 7, stress levels soared, with PTSD prevalence reaching 22.9-36.0% at 7-8 weeks post-event, and remaining exceptionally high at 15.9-24.7% after 7 months. We found a strong correlation between increased PTSD risk and extensive media consumption. Smartwatch data revealed early PTSD markers such as pronounced increases in stress levels and declines in mood, physical activity, and sleep quality within the first week. This study underscores the significant impact of media exposure on PTSD development and the value of continuous physiological monitoring for early detection of PTSD after mass traumatic events.

3 - Advancing Machine Learning-Based Job Recommender Systems: Human Skills-Based Approach**Hila Chalutz-Ben Gal, Bar-Ilan University, Ramat Gan, Israel**

In recent years the widespread adoption of machine learning-based job recommender systems became a major component in predictive human resources analytics. However, recent workforce disruptions such as freelancing, gig, and flexible work have resulted in the emerging need to adopt new models and tools to support and implement state-of-the-art job recommender systems. To address contemporary workforce disruptions, we utilize the Person-Skill Fit framework to present a robust analysis of the machine learning based job recommender systems literature. We adopt a scoping review method to systematically analyze seventy-three papers based on predetermined inclusion and exclusion criteria. Our results indicate that job recommender systems may benefit from a skill-based approach and should therefore be adopted to address new workforce challenges. The research main contribution is a decision support tool for implementing machine learning-based job recommender systems and upskilling for improved human capital management. KEYWORDS: Job Recommender Systems; Human Resources Analytics; Machine Learning; Iterative Scoping Review.

4 - Continuous Monitoring and Early Detection of Health and Well-Being Conditions Using Smartwatches: a Systematic Review of AI Advances**Liora Shmueli, Bar-Ilan University, Tel Aviv, Israel, Maya Reiner, Erez Shmueli**

This systematic review investigates the role of smartwatches in the continuous monitoring and early detection of health and well-being conditions, with a focus on recent advances in artificial intelligence (AI). Continuous monitoring technologies are gaining prominence as they offer the potential for early diagnosis and timely intervention, especially when combined with machine learning and deep learning techniques. As off-the-shelf smartwatches become increasingly popular beyond fitness tracking, their integration into healthcare contexts has expanded considerably over the last few years.

This review follows the PRISMA guidelines and based on the Cochrane Handbook for Systematic Reviews of Interventions. A comprehensive search was conducted in early 2025 across PubMed, Scopus, and IEEE databases. The review covers peer-reviewed studies published between 2014 and 2024. Studies were eligible if they provided quantitative evidence

on the accuracy or clinical utility of biometric data collected via commercially available smartwatches, combined with machine learning and deep learning techniques. Over 6,000 articles were screened, resulting in a final sample of studies that met rigorous inclusion criteria.

Our findings emphasize the positioning of AI-enhanced smartwatches as key instruments in monitoring and detecting health and well-being conditions, as well as offering predictive insights that can improve patient outcomes and support personalized healthcare.

5 - Exploring the Potential of Large Language Models for Vestibular Rehabilitation Education: a Comparative Evaluation of ChatGPT, Google Gemini, and Clinicians**Yael Arbel, Bar-Ilan University, Ramat-Gan, Israel, Yoav Gimmon, Liora Shmueli**

This study aimed to compare the performance of two large language models, ChatGPT and Google Gemini, against experienced physical therapists and students in responding to multiple-choice questions related to vestibular rehabilitation (VR). The study further aimed to assess the accuracy of ChatGPT's responses by board-certified otoneurologists.

Thirty physical therapist professionals experienced with VR and Thirty physical therapy students completed a 20-item multiple-choice questions Vestibular Knowledge Test (VKT), divided into three categories: Clinical Knowledge, Basic Clinical Practice, and Clinical Reasoning. ChatGPT and Gemini were given the same test. Three board-certified otoneurologists independently evaluated the accuracy of each response using a 4-level scale, ranging from comprehensive to completely incorrect.

ChatGPT3.5 outperformed Google Gemini with a 70% score on the VKT test, while Gemini scored 60%. Both excelled in Clinical Knowledge scoring 100% but struggled in Clinical Reasoning (ChatGPT: 50%, Gemini: 25%). According to three otoneurologic experts, ChatGPT's accuracy was considered "comprehensive" in 45% of VKT, while 25% were found to be completely incorrect. ChatGPT provided "comprehensive" responses in 50% of Clinical Knowledge and Basic Clinical Practice questions, but only 25% in Clinical Reasoning. To examine model improvements, updated versions of both LLMs were also evaluated on the same test; results are under analysis and will be presented.

Findings highlight that while LLMs show promise in factual knowledge, their clinical reasoning remains limited. Use of LLMs in clinical education should be approached cautiously, with careful question formulation and expert validation. Integrating LLMs with clinical expertise and guidelines may enhance learning while mitigating risks

6 - A Bi-Level Optimization Decision Support Tool for Maximizing Company Profit Through Employee Retention**Iris Forma, Tel-Aviv Afeka Academic College of Engineering, School of Industrial Engineering and Management, Tel Aviv - Yafo, Israel, Inbal Singer, Gonen Singer, Hila Chalutz Ben-Gal**

The rise of flexible work greatly impacts how firms manage and retain employees. This research introduces two bi-level mathematical models to help companies improve workforce retention by integrating flexibility and attractive benefits. The upper level focuses on maximizing firm profit, while the lower level aims to minimize the gap between desired salary and total benefits for employees. Model I assumes fixed salaries for each work plan and job type, whereas Model II allows for fully flexible salary determination. Both models provide exact solutions using a mixed-integer formulation. Experimental analysis shows that in some scenarios, Model II leads to significantly

higher firm profit than Model I. However, when work plan flexibility is similar across markets and firms, the profit difference is minimal. These results highlight the practical relevance of our solutions. This study makes three important contributions. First, from a methodological perspective, it introduces a novel bi-level modeling approach that can proactively enhance employee retention and, consequently, decrease employee turnover. Second, from a practical perspective, it offers concrete solutions to new workforce challenges by providing a decision support tool capable of optimizing employee retention through the creation of customizable and flexible work plans and salaries, specifically designed to meet diverse employee needs while simultaneously maximizing the company's profit. Third, the study highlights the critical importance of incorporating managerial insight into the model's parameters and effectively utilizing experimental analysis as a vital decision support mechanism to achieve optimal firm profitability.

MB45

Building B Level 3 B306

Technology and Society

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Yanhan (Savannah) Tang, SMU Cox School of Business, 6214 Bishop Blvd, Dallas, TX, 75275, United States

1 - The Effect of Corporate Participation on Volunteer Contributions in Online Communities

Eunae Yoo, Indiana University, Bloomington, IN, United States, Yuqing (Aric) Cheng, Abhishek Nagaraj, Elliot Rabinovich

Online communities often rely on volunteer contributions to generate content, but increasingly attract direct corporate participation as they mature. We examine how corporate contributions affect volunteer engagement in these platforms, addressing a gap in the literature where existing theories offer conflicting predictions. Using OpenStreetMap data from Brazil (2017-2022), we employ an event study methodology to analyze volunteer mapping activities before and after corporate entry across different regions. Our findings reveal that corporate participation significantly increases volunteer contributions rather than crowding them out. This positive effect is particularly pronounced when corporate contributors engage broadly with the platform rather than focusing narrowly on certain regions. We find that such corporate involvement motivates dedicated volunteers to not only contribute more but also redirect their efforts toward previously neglected and less mature areas of the map. These results suggest that strategic corporate participation can enhance overall platform quality by complementing volunteer efforts. Our work contributes to understanding the complex dynamics between different contributor types in online communities and offers insights for platform governance that effectively balances volunteer and corporate participation.

2 - Analyzing Social Media Discourse on Child Protective Services to Enhance Timely Support

Yanhan (Savannah) Tang, SMU Cox School of Business, Dallas, TX, United States

The rise and popularity of social media platforms like Reddit have created a valuable opportunity to study public discourse around child protective services (CPS). This project leverages large language models to analyze Reddit posts and comments related to CPS, assess sentiment-topic alignment, and identify post characteristics associated with receiving timely support. Based on these results and insights, we offer recommendations to help users improve the quality and effectiveness of their posts so that they are more likely to receive timely and helpful responses from the community.

3 - Theory-Guided Representational Richness for LLM-Based Assessment

Xinyuan Zhang, UNIVERSITY OF NOTRE DAME, SOUTH BEND, IN, United States, Junhui Cai, Ahmed Abbasi

With the current rapid development in generative AI, pretrained models, including Large Language Models (LLMs), are increasingly being for assisting humans in evaluating both human and generative outputs. This increasing LLM usage presents several opportunities for improving and expanding on current NLP studies. LLMs are capable of grasping richer information across a large lexicon domain, and among large volume inputs. However, further processing the multi-dimensional and large-scale information from LLMs can be computationally challenging, while over-simplified techniques fail to pertain granular and valuable signals. We propose a theory-guided framework that creates parsimonious representations of multi-dimensional information from LLM-based assessments with downstream impact. We use tensor-based representations and extract meaningful patterns, that effectively address interplay between tokens and parallel representations for text. We compare our representations with state of the art methods, and demonstrate downstream values in various practical tasks through case studies across several testbeds.

4 - Can AI-Generated Stories Heal the Heart?

Jiayi Xu, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Yifan Yu, Wendao Xue, Junyu Yao, Qiaoxiang Hu

Stories are central to how people understand the world and themselves. In mental health, stories are known to support healing. Immersion in resonant stories encourages emotional processing, perspective-taking, and self-reflection, while reducing psychological resistance to new insights. However, stories created by humans are limited and people may not always have access to high-quality and personally relevant stories. The rise of generative AI (GenAI) offers new possibilities: by rapidly producing tailored, empathetic stories, GenAI enables people to engage with and benefit from stories that feel personally meaningful. This study aims to understand a novel conversational healing mode in which people converse with a chatbot based on personalized, GenAI-generated stories. We will conduct an experiment to understand two fundamental questions: (1) Is this story-based healing effective? (2) How do people benefit from this healing mode? Our research expects to contribute empirical evidence on the healing potential of GenAI-powered storytelling and to uncover the behavioral mechanisms that underlie its impact.

5 - Optimizing the Ineffable: Generative Policy Learning for Human-Centered Decision-Making

Zekai Fan, Carnegie Mellon University, Pittsburgh, PA, United States, Michael Lingzhi Li, Shixiang (Woody) Zhu

Algorithmic decision-making is widely adopted in high-stakes applications affecting our daily lives but often requires human decision-makers to exercise their discretion within the process to ensure alignment. Explicitly modeling human values and preferences is challenging when tacit knowledge is difficult to formalize, as Michael Polanyi observed, "We can know more than we can tell." To address this challenge, we propose generative near-optimal policy learning (GenNOP). Our framework leverages a conditional generative model to reliably produce diverse, near-optimal, and potentially high-dimensional stochastic policies. Our approach involves a re-weighting scheme for training generative models according to the estimated probability that each training sample is near-optimal. Under our framework, decision-making algorithms focus on a primary, measurable objective, while human decision-makers apply their tacit knowledge to evaluate the generated decisions, rather than developing explicit specifications for the ineffable, human-centered objective. Through extensive synthetic and real-world experiments, we demonstrate the effectiveness of our method.

MB46

Building B Level 3 B306

Entrepreneurial and Analytics Path for Decisions: Policy, Energy, Water, and Innovations

Invited Session

Decision Analysis Society

Chair: Christian Hernandez-Negron, California State University, Amherst, United States

1 - Discrete Event Simulation for Nuclear Waste Disposition

Veronica White, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Alexis Andaverde, Hiram Moya

Nuclear waste disposal is a growing issue, with significant backlogs of nuclear waste at many national laboratories and research centers dealing in medical research, clean energy, and other nuclear activities. This backlog prevents future use of nuclear materials, risks environmental impacts, and complicates federal and state regulatory

compliance, making disposal costly. Rising interest in nuclear energy will further inflate the issues of backlogged nuclear waste. Therefore, effectively managing and disposing of nuclear waste is critical to invest in future nuclear solutions. Simulation approaches can provide insight into effective resource allocation and waste processing operations to stabilize productivity and cut the backlog of nuclear waste. Therefore, we created a discrete event simulation in ExtendSim to model a nuclear waste disposition process, which is a process within nuclear waste disposal. We inform the simulation by evaluating a case study of the Savannah River National Laboratory (SRNL), which is a United States nuclear waste disposal facility, using data and insights from SRNL staff.

2 - Watts and Drops: Renewable-Aware Co-Optimization of Power and Water in Desalination Plants

Ahmed Alahmed, King Fahd University of Petroleum and Minerals, Dharhan, Saudi Arabia, Audun Botterud, Saurabh Amin

As water and energy systems become increasingly interdependent, there is a growing need for integrated optimization frameworks that enable cost-effective and grid-responsive operation of critical infrastructure such as water desalination plants (WDP). We develop a mathematical framework to co-optimize water and electricity scheduling in a profit-maximizing WDP that operates both thermal- and membrane-based technologies. The price-taking WDP sells desalinated water to a water utility at a given price and engages in bi-directional electricity transactions with the grid, purchasing or exporting power based on its net electricity demand. We show that the optimal scheduling policy depends on the plant's internal renewable generation and follows a simple threshold structure. Under the optimal policy, thermal-based water output decreases monotonically with renewable availability, while membrane-based output increases monotonically. We characterize the structure and intuition behind the threshold policy and examine key special properties.

3 - Modelling to Generate Continuous Alternatives: Enabling Real-TIME Feasible Portfolio Generation in Convex Planning Models

Michael Lau, Princeton University ZERO Lab, Princeton, NJ, United States, Neha Patankar, Jesse Jenkins

Modelling to Generate Alternatives (MGA) provides an optimization method for capturing many near-cost-optimal system configurations, and can provide insights into the tradeoffs between objectives and flexibility available in the system. However, MGA is currently limited in interactive applicability to these problems due to a lack of methods for allowing users to explore near-optimal feasible spaces. In this work we describe Modelling to Generate Continuous Alternatives (MGCA), a novel post-processing algorithm for convex planning problems which enables users to rapidly generate new interior solutions, incorporate new constraints, and solve within the space with convex objectives. MGCA begins with a dimensionality reduction to capacity decisions and metric values. We then take advantage of convex combinations to generate interior points by allowing user weight specification and encoding convex combinations in an optimization problem with user-defined additional constraints and objective. Dimensionality reduction enables this problem to solve in tenths of a second, suitable for analysis in interactive settings. We discuss the interpolation of capacity and operational metric values, finding capacity metrics can be perfectly interpolated while operational metrics remain within the feasible range of the points used to create them. We demonstrate interpolated solutions can be exported and re-solved with an economic dispatch model to provide operational metric values consistent with least-cost decision-making and show interpolated metric values are generally within 10% of the optimal value.

4 - Multi-Vector Optimization Framework to Support Equitable Building Decarbonization in Urban Environments

Jun Wen Law, Massachusetts Institute of Technology, Cambridge, MA, United States, Dharik Mallapragada

Regional building decarbonization efforts in cold climates emphasize electrification coupled with expansion of low-carbon electricity supply. This transition requires substantial investments across the entire electricity supply chain—generation, transmission, distribution, and end-use—while also addressing the role of existing natural gas infrastructure currently serving both building and power sectors. This study presents a mixed-integer optimization framework for joint electricity and natural gas infrastructure planning to support building decarbonization. Unlike prior studies that generally emphasize bulk energy infrastructure aspects of this problem, our approach spans both transmission and distribution networks, offering a comprehensive evaluation of the system impacts of alternative building decarbonization strategies. The model evaluates the least-cost investment and operation of the integrated energy system to meet heating and non-heating demands, while accounting for: a) temporal and spatial variability in demand, b) network operations, c) end-use technology adoption, d) policy drivers such as emissions targets and distributional equity, and e) a wide range of low-carbon technologies including clean generation, storage, low-carbon fuels, and district heating. The value of the framework is demonstrated by evaluating a case study developed using data from New York City. Preliminary results evaluating alternative pathways to achieve mid-century decarbonization targets show that hybrid pathways—such as electrification with low-carbon fuels—as opposed to full electrification can reduce the urgency for rapid electrification, preserve existing infrastructure, and prevent premature asset retirements and costly upgrades. The study also highlights the value of coordinated investment planning across both bulk and distribution energy systems to support cost-effective and equitable decarbonization.

MB47

Building B Level 3 B307

Optimization Topics V

Contributed Session

Contributed

Chair: Michael Geisecke, University of Missouri

1 - Optimizing HVAC Control for Energy Efficiency and Occupants' Health and Productivity

Ziwei Liu, University of Tennessee, Knoxville, TN, United States, Mingzhou Jin, Zhihong Pang

Buildings significantly contribute to global energy consumption, with Heating, Ventilation, and Air Conditioning (HVAC) systems accounting for approximately 38% of commercial building energy usage. As people spend roughly 90% of their time indoors, maintaining optimal Indoor Environmental Quality (IEQ), including thermal comfort, volatile organic compound (VOC) levels, and CO₂ concentration, is critical for occupants' health and productivity. Thus, balancing energy efficiency and IEQ constitutes a core optimization challenge in HVAC control in buildings.

This paper develops a physics-based optimization model integrating thermal dynamics, indoor air quality, pollutant dispersion, and HVAC energy consumption. To handle inherent nonlinearity and nonconvexity, this study introduces a solution-oriented piecewise McCormick relaxation for bilinear terms, along with a feasibility adjustment algorithm to ensure the relaxed formulation yields a feasible solution to the original constraints. Four methods (Baseline, Gurobi, Partial Approximation, and Full Approximation) are benchmarked over 48 scenarios involving varying temperatures and outdoor air pollution levels, each scenario executed over 10 instances. Performance metrics include average CPU time, solution deviations, and statistical significance (P-values). Results illustrate critical trade-offs among energy efficiency, feasibility, and computational cost. Specifically, the Full Approximation approach achieves superior energy efficiency at high computational cost, whereas Baseline, while faster, frequently yields infeasible solutions.

Ultimately, our proposed optimization framework provides a comprehensive and practically implementable strategy for improving building sustainability and energy efficiency without compromising indoor environmental quality for occupants' health and productivity. In addition, the benchmarking analysis provides quantitative guidance for selecting solution methods that best balance energy savings, feasibility, and computational burden.

2 - On Solution Structure and Stability of Optimality-Based Clustering

Soroush Akbarijokar, University of Pittsburgh, Pittsburgh, PA, United States, Taewoo Lee

We develop a new clustering problem where data points represent decisions made by solving (partially known) optimization problems and are clustered based on their latent decision-making preferences, which we term the optimality-based clustering problem. While the problem is first modeled as a large-scale mixed-integer program, we elucidate its connection to facility location problems, which we exploit to derive a more efficient reformulation. We further propose a stability-driven variant of the problem to mitigate the impact of data noise and uncertainty on the clustering solution. We discuss practical implications of this framework in healthcare applications including eliciting group diet preferences and clustering clinical pathways based on underlying clinical preferences.

3 - Facility Layout Adaptation Optimization: a Matheuristic Approach

Mingze Li, Georgia Institute of Technology, Atlanta, GA, United States, Jihye Jung, Benoit Montreuil

This study introduces a novel Facility Layout Problem variant, FLP with Base Layouts (FLP-BL), addressing the challenge of automating the design of consistent industrial layouts across multiple facilities. Current industry practice relies heavily on manual, heuristic-based design, especially for complex layouts with integrated material handling networks. In fast-evolving, geographically distributed industries, typical of today's market, companies may operate multiple facilities in different locations that share identical set of resources. Each new facility requires a distinct layout that fits the new building while maintaining consistency across locations to enhance operational familiarity, an aspect previously overlooked in academic research. We propose a three-stage mixed-integer programming (MIP) based matheuristic model to automate facility layout design. The objective is to minimize material handling flow cost while preserving the core structure of expert-designed base layouts. Constraints ensure stations are within building boundaries, non-overlapping, and do not overlap with aisles, while maintaining aisle network connectivity for material handling origin and destination pairs. A real-world case study demonstrates that our model achieves lower flow costs within one-third of computational time compared to directly solving the base MIP model for 24 hours. Furthermore, we show that with minor changes, our matheuristic model can also be used to solve the general FLP from scratch. This research offers a practical and efficient approach to automated facility layout design for modern industrial needs.

4 - Qualitative-Metric Clustering: Minimal Clique Cover with Minimal Edge-Cuts

Yuxuan Ren, Georgia Tech, Atlanta, GA, United States, Shijie Deng

This study presents a qualitative-metric clustering framework that concurrently balances three objectives—high within-cluster cohesion, low between-cluster similarity, and a parsimonious number of clusters. We tackle the ensuing combinatorial problem with a branch-and-price scheme: the restricted master problem is modeled as a weighted set-covering formulation, and the pricing sub-problem is expressed as a maximum-weight clique search. A destroy-and-repair heuristic is embedded in the pricing step to generate high-quality columns early in the search. The methodology is validated on Financial Transmission Rights (FTR) trading data, where profitable FTR paths are identified by clustering on Shift Factor (SF) matrix, demonstrating both its practical relevance and its effectiveness.

5 - Decision Support for Integrating Remanufacturing and Recycling Facilities into Manufacturing Plants Under Sourcing Cost Uncertainties and Product Lifecycle Considerations

Mohammad Ahnaf Sadat, Iowa State University, Ames, IA, United States

In this study, we examine the strategic decision-making involved in integrating remanufacturing and recycling facilities into existing manufacturing plants, using turbine blade production for gas turbines as a representative case study. Our analysis emphasizes uncertainties associated with sourcing costs, particularly manufacturing and remanufacturing expenses, which are modeled as stochastic dynamic variables. Additionally, the analysis accounts for variations in product demand throughout the product lifecycle and treats the returned products' quality during remanufacturing as deterministic dynamic variables. To quantify the benefits of investment flexibility, we employ Real Options Analysis, utilizing Monte Carlo simulations. Our objectives include identifying optimal conditions for facility integration, determining the strategic value derived from investment flexibility under uncertain economic conditions, and investigating how key factors, such as cost variabilities, product lifecycle dynamics, and the quality of returned items, impact the overall decision-making process.

6 - Optimal Hurricane Evacuation Routing with Private and Public Vehicles: a Joint Chance-Constrained Stochastic Programming Model

Michael Geisecke, University of Missouri, Columbia, MO, United States, Hyeon Suk Na, James Noble

Efficient evacuation planning is crucial for mitigating the adverse effects of major hurricanes, especially in urban areas where a substantial portion of the population relies on public transportation on a daily basis. The evaluation of evacuations from major hurricanes such as Hurricane Idalia, Ian, and Barry has revealed critical gaps in evacuation planning, with inadequate public transportation contributing to loss of life. However, the role of public transportation remains underrepresented in existing hurricane evacuation studies. To address this gap, we develop a stochastic programming model that integrates the cell transmission model with chance constraints, which provides a robust framework for capturing uncertainties in road capacity and assessing their impact on the evacuation of public vehicles. The objective of the proposed joint chance-constrained programming (JCCP) model is to support emergency management decision-making by minimizing the total cumulative time that evacuees spend within the evacuation transportation network. We reformulate the JCCP model using the sample average approximation method, which employs Monte Carlo simulations to approximate the distributions of uncertain parameters. This reformulation results in a mixed-integer linear programming model, offering a more computationally efficient alternative to solving the original stochastic optimization problem. The applicability of the proposed model is validated through a case study focused on the Miami metropolitan area in Florida. The experimental results confirm our model's effectiveness in generating optimal evacuation routes across various hurricane scenarios. Additionally, our model estimates the minimum number of buses required for evacuations and evaluates alternative bus routing policies to reduce traffic congestion during evacuations.

MB48

Building B Level 3 B308

Disinformation and Impact on Infrastructure Systems

Invited Session

Decision Analysis Society

Chair: Kash Barker, University of Oklahoma, Norman, OK, United States

1 - Modeling the Impact of Disinformation on Infrastructure Networks with Causal Failures Across Interconnected Layers

Alice Nanyanzi, University of Oklahoma, Norman, OK, United States, Kash Barker, Sridhar Radhakrishnan

The widespread dissemination of disinformation poses a serious threat to society, especially when used to disrupt the performance of critical networks. We model this threat using a two-layer network consisting of an information layer (A), where dis/information spreads, and a physical layer (B), where the physical network operates. For example, false messages about electricity rate hikes in layer A can alter consumer behavior, increase demand, and potentially overload power lines, leading to blackouts in layer B. These disruptions reveal how vulnerabilities in one layer can propagate to another. To capture the dynamics of disinformation spread in layer A, we develop an adaptive influence model that reflects how decision-making and heterogeneous susceptibility evolve during disinformation campaigns. To examine the interdependencies between layers, we present a novel approach that focuses on direct cause-and-effect relationships, where the likelihood that a node in A causes a path failure in B is governed by stochastic processes. We study this by formulating it as problem that determines the minimum number of stochastic causal failures in A required to significantly disrupt the ability of B to meet demand or in the worst case cause widespread service interruptions in B. Illustrating with data from social media and power grid networks, we demonstrate how disinformation in layer A can activate causal pathways that result in stochastic failures in infrastructure layer B. This framework provides insight into the cascading risks that arise from the interaction between information and physical systems.

2 - Interdicting Disinformation Spread to Mitigate Impacts on Public Transit Networks

Daniel Cabrera, University of Oklahoma, Norman, OK, United States, Andrés Gonzalez, Kash Barker

The growing influence of online information has increased the exposure of critical infrastructure systems to the risks of disinformation. This study examines how false messages about transit station closures can disrupt public transportation systems by altering user behavior. We propose a two-layer modeling framework that captures both the spread of disinformation and its impact on transit operations. The information layer models user susceptibility based on personality traits and simulates message propagation through an information diffusion process. The physical layer represents the transit network as a multi-commodity flow model, capturing user movement under different information awareness scenarios. These layers are integrated to assess how disinformation alters traffic patterns and degrades system performance. We apply this framework to a case study of the New York City subway system. Our results demonstrate that disinformation can significantly increase travel times and system congestion. However, targeted protection strategies—such as informing a small, strategically chosen subset of users—can mitigate these effects and improve resilience. This work highlights the importance of incorporating behavioral and network dynamics into infrastructure protection models. It provides a novel operational research approach to evaluate and counter the systemic risks posed by disinformation.

3 - Understanding the Role of Emotion in the Spread and Impact of Disinformation

Elena Bessarabova, University of Oklahoma, NORMAN, OK, United States, Matthew Weber, Casey Randazzo

Understanding the role of emotion in the spread and impact of disinformation is essential. Emotions such as fear, anger, and anxiety can override rational judgment and decision making, making individuals more susceptible to disinformation and conspiratorial content. Disinformation often strategically leverages these emotions to generate urgency, foster distrust in institutions, and motivate maladaptive decision-making such as the rejection of public health measures. Disinformation does not merely disseminate falsehoods—it deliberately targets emotional vulnerabilities to enhance engagement. Emotionally charged content spreads more rapidly, leaves stronger cognitive impressions, and can distort memory and perceived credibility through mechanisms such as the availability heuristic. In the case of conspiratorial ideation, once individuals become emotionally invested in conspiracy theories, they exhibit greater resistance to corrective information, underscoring the need for early intervention. This presentation will discuss recent empirical findings on the role of emotions such as fear in disinformation processing, including its influence on disinformation spread and resistance to conspiratorial beliefs. We also discuss intervention strategies that aim to reduce emotional susceptibility to disinformation, particularly in crisis contexts where emotional vulnerability is heightened. These findings have important implications for designing more effective counter-disinformation efforts and public communication strategies.

4 - Modeling Disinformation as Wave Dynamics: a Quantum Graph Framework for Belief Evolution

Sridhar Radhakrishnan, University of Oklahoma, Norman, OK, United States, Kash Barker

Disinformation does not spread as a simple contagion but rather evolves as a complex, continuous process shaped by repeated exposure, reinforcement, and local social network structure. Classical models based on epidemic dynamics fall short in capturing this complexity, as they rely on discrete state transitions (e.g., susceptible, infected, recovered) that assume binary belief states. To address such limitations, we introduce a novel modeling framework based on quantum graphs, which naturally encode belief as a continuous, dynamic process akin to wave propagation. We begin with an information graph (IG), where nodes represent individuals and edges represent influence links. From this, we construct a quantum graph (QG) as the incidence graph of IG, where each node in QG corresponds to an edge in IG, and each edge in QG connects QG nodes that share a common IG node. In this formulation, the wave function defined on each edge of QG models the evolving belief state of a node in IG, capturing how individuals gradually adopt, resist, or amplify disinformation based on their interactions. We model belief evolution using wave functions governed by time-dependent partial differential equations, such as the Schrödinger or wave equation. These wave functions enable the modeling of continuous and temporally sensitive belief change, including interference from multiple sources, reinforcement through repetition, and attenuation due to skepticism or conflicting signals. The core strength of this quantum graph framework lies in the ability to encode belief dynamics through the deliberate design of wave functions and boundary conditions.

MB49

Building B Level 3 B309

Sustainability and Resilience in Supply Chains

Invited Session

MSOM: Supply Chain

Chair: Shi Chen, University of Washington Foster School of Business, Seattle, WA, United States

Co-Chair: Xiaoxuan Hou, University of Washington, Seattle, WA, United States

1 - Greenwashing Under Competition

Soraya Fatehi, The University of Texas at Dallas, Richardson, TX, United States, Anyan Qi, Liqun Wei, Jianxiong Zhang

Growing consumer awareness of corporate social responsibility (CSR) has motivated firms to invest in CSR initiatives to gain a competitive edge. However, a phenomenon known as greenwashing has emerged, whereby firms exploit observable CSR activities and advertising solely as a marketing tactic. We develop a game-theoretic model with two types of firms: a socially responsible firm that intrinsically values CSR and a profit-maximizing firm that may greenwash. Firms can invest in observable and/or unobservable CSR activities as well as CSR advertising. Consumers are socially minded but face limited information about the firms' actual CSR type, inferring the type through observable CSR investment and advertising. We examine how the advertising influence and competition intensity affect equilibrium strategies and social welfare. Our findings show that when advertising influence is high, the profit-maximizing firm engages in greenwashing; when it is moderate, the socially responsible firm overinvests in CSR to deter imitation; and when it is low, the two firm types naturally separate. Notably, under high transparency, stronger advertising influence can increase both overall CSR activity and social welfare. Moreover, when transparency is low, the profit-maximizing firm strongly prefers greenwashing; in such settings, if competition intensity is low, greenwashing persists, whereas if competition intensity is high, the socially responsible firm responds by over-investing to prevent greenwashing. Greenwashing produces both negative and positive outcomes. Although it erodes consumer surplus, it can spur higher CSR activity in a competitive environment, enhancing social welfare.

2 - Collective Renewable Power Purchase Agreement: Negotiation and Risk-Sharing Mechanisms

Xiaoxuan Hou, University of Washington, Seattle, WA, United States, Shi Chen

Background: Over 400 companies spanning more than 175 markets worldwide have committed to using 100% renewable electricity. But achieving these goals can be challenging. In the past decade, long-term power purchase agreement (PPA) has gained popularity among large corporations like Google and Amazon (buyers) as the main vehicle to source supply from renewable energy developers (sellers). Under a PPA, the buyer commits to purchasing renewable energy from the seller at a contract price for a predetermined length of time. Yet signing a PPA remains difficult for small-to-medium-sized corporate buyers, who may not have the expertise or lack sufficient negotiation power on their own. As a result, collective PPAs have emerged as a potential solution.

Method: We compare the performance of three widely used negotiation mechanisms for buyers to sign PPA contracts: separate negotiation, agent-based negotiation, and joint negotiation. For each mechanism, we devise a two-stage problem involving Nash-bargaining based negotiation, followed by post-contract operational optimization.

Results: Our analysis reveals the impact of different mechanisms on the renewable project size in equilibrium as well as different players' preference. Assuming risk neutrality, we find that the three negotiation mechanisms lead to identical equilibrium project sizes. However, conflicts of preference for different negotiation mechanisms exist among parties. Further, buyers' perception of risks due to market uncertainty and REC market prices can harm the achievement of the sustainable goals.

3 - Building Supply Chain Resilience and Agility Using ML & AI

Vinayak Deshpande, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Naren Agrawal, Morris Cohen

The global pandemic and other recent geopolitical events have highlighted the importance of supply chain resilience and agility. Unfortunately, many companies have failed to develop strategies that can deliver these capabilities. In particular, the most promising enablers, i.e., artificial intelligence and machine learning, cloud computing, and big data, have fallen short. In this presentation, we describe Optimal Machine Learning, a new paradigm that we have developed, that leverages the capabilities of advanced technology. We illustrate the power of this new approach for supply chain planning by reporting on two successful implementations in complex global supply chain environments. We then discuss a framework to guide efforts to design effective strategies for supply chain resilience and agility.

4 - Building Supply Chain Resilience Through Vertical Control

Ali Kaan Tuna, Tilburg University, Tilburg, Netherlands, Nitin Bakshi, Robert Swinney

We study the value of two types of vertical control—vertical integration and direct sourcing—for increasing resilience towards random raw materials shortages. We use a three-tier supply chain model in which a Tier 0 buyer sources a critical component from a Tier 1 supplier, who sources raw materials from disruption-prone Tier 2

suppliers. With vertical integration, the buyer purchases its Tier 1 supplier, taking control of its sourcing decisions. With direct sourcing, the buyer purchases raw materials from Tier 2 and sells to Tier 1. We find that either form of vertical control is most valuable when the likelihood of disruptions is moderate, with vertical integration preferred over direct sourcing for more severe disruptions. We next consider whether penalty contracts—in which the buyer charges the supplier a penalty for each unit the supplier fails to deliver—can replicate the outcome of vertical control, and find that while this is possible, the necessary non-delivery penalty may be too high for the supplier to accept, particularly when disruptions are of moderate likelihood, implying vertical control can be valuable even if penalty contracts are possible. Finally, we explore whether multi-sourcing in Tier 1 can eliminate the need for vertical control and find that while it can reduce the need for vertical control, vertical control is still optimal when disruption risk is moderate. Together, our results show that, under specific circumstances, vertical control can be a valuable addition to the portfolio of strategies firms use to mitigate disruption risk.

MB50

Building B Level 3 B310

Frontier Research on Innovative Service Operations

Invited Session

MSOM: Service Operations

Chair: Ming Hu, Rotman School of Management, University of Toronto, Toronto, ON, Canada

Co-Chair: Zhoupeng (Jack) Zhang, University at Buffalo School of Management, The State University of New York, Buffalo, NY, United States

1 - Congestion in the Age of Uber: Medallion Management and Taxi Hailing Technology

Kaitlin Daniels, Wisconsin School of Business, University of Wisconsin, Madison, WI, United States

The introduction of Transportation Network Companies (TNCs) into ride-hailing markets poses a challenge for congestion-sensitive regulators. Many large cities have historically been served exclusively by street-hailing taxis, which, to operate, require one of a limited number of medallions issued by a regulator. TNCs differ from taxis in two important ways: (i) the regulator does not control their supply of self-scheduling drivers, and (ii) they use a more efficient hailing technology (“e-hailing”). We study how the use of two existing regulatory powers – issuing medallions and restricting taxi hailing technology – can address congestion caused by the TNC. To minimize congestion, the regulator decides how many medallions to issue and whether to improve taxi hailing by allowing taxis to e-hail. The regulator’s optimization is influenced by the equilibrium behavior of TNC drivers, who choose whether to drive, and passengers, who choose to patronize taxis or the TNC and for whom the regulator must ensure sufficiently fast pickups. In many cities, the regulator’s existing powers are sufficient to mitigate congestion caused by the TNC. This is accomplished by improving taxi hailing and reducing the number of medallions in circulation. However, our results caution that regulators should be wary of the seeming efficiency of improving taxi hailing, which may increase overall congestion. This work provides concrete recommendations to regulators, identifying when to increase or decrease the number of medallions in circulation and/or improve taxi hailing in response to the presence of a TNC.

2 - Blockchain Traceability and Customer Satisfaction: Evidence from E-Commerce

Farzad Fathi, University of Maryland, College Park, MD, United States, Jane Jiang, Jiding Zhang

Blockchain technology offers enhanced traceability and transparency, making it particularly valuable for products susceptible to counterfeiting or those for which provenance is critical. Leveraging data from an e-commerce platform that integrates blockchain-based traceability, we empirically examine the impact of this technology on customer satisfaction. To complement our empirical findings, we develop a game-theoretical model that captures the strategic role of information disclosure enabled by blockchain traceability.

3 - Order Ahead for Pickup: Promise Or Peril?

Yunan Liu, Amazon, New York, NY, United States

Mobile technologies increasingly allow customers to order ahead at quick-service restaurants, potentially reducing delay and increasing throughput. We develop a queueing-game-theoretic model to study a hybrid system where some customers order ahead and others order onsite. Surprisingly, we find that accepting all orders as they arrive and making them irrevocable can lead to lower throughput in the hybrid system than in a pure onsite system. This inefficiency persists even when queue-length information is shared with remote customers. To address this, we examine two mitigation strategies. First, allowing remote customers to cancel orders not ready upon arrival can restore the throughput advantage of ordering ahead, although it may still underperform the non-cancellation scheme. Second, implementing a rejection policy—where new remote orders are blocked once outstanding orders exceed a threshold—can improve throughput when optimally calibrated, outperforming both the pure-onsite and non-rejection hybrid schemes, though not always the cancellation scheme. We then propose an integrated mechanism that jointly optimizes both rejection and cancellation thresholds, subject to consumers’ individual rationality. This mechanism encompasses all previously considered schemes and offers the best flexibility and performance. Our findings reveal the unintended consequences of naive order-ahead implementations and provide prescriptive guidance for managing hybrid service systems to enhance throughput while maintaining service quality.

4 - Freelancers OR Employees? a Field Experiment with a Food Delivery Platform

Wee Kiat Lee, Nanyang Technological University, Singapore, Singapore, Yao Cui, Qi Li

The ongoing debate surrounding the regulation of gig platforms to classify service providers as employees rather than freelancers has generated significant industry discussions. However, the potential impact of such a transition on the operational efficiency of gig platforms is largely unknown. In this paper, we adopt an operational perspective to examine this issue. We collaborate with a food delivery platform and implement a field experiment in which orders are randomly routed to freelancers and employees. Using the experimental results, we compare the job performance between freelancers and employees, and derive insights into how a platform should adjust its operational policies when it changes the recruitment status of service providers.

5 - Implications of Worker Classification in on-Demand Economy

Zhoupeng (Jack) Zhang, University at Buffalo School of Management, The State University of New York, Buffalo, NY, United States

How should gig workers be classified, as employees or as contractors? We study this policy question by highlighting the difference between full- and part-time gig workers and focusing on the welfare of full-timers, who do gigs to earn primary incomes. We develop a multi-period game-theoretic queueing model to study market outcomes in different worker classification scenarios. We show that being classified as independent contractors is not always harmful to full-timers; they can end up with high surpluses thanks to a positive incentives de-pooling effect of part-timers’ market entry. As such, when all gig workers are reclassified as employees (as per, e.g., Assembly Bill No. 5 in California) and part-timers exit the market, full-timers can be undercut (underhired or underpaid) by their platform employer and end up with lower surpluses. When all are reclassified as “contractors+,” a UK practice that provides incomplete employee benefits but still gives workers flexibility, additional part-timers incentivized may overcrowd the platform and hinder the improvement in full-timers’ welfare. We propose two policy alternatives: (1) a Hybrid Mode that classifies full-timers as employees but part-timers as contractors, and (2) a Priority Mode that prioritizes matching full-timers with customers over part-timers. We verify that these differentiated schemes can benefit full-timers as well as other stakeholders in the market. We calibrate the parameters in our model using real-world data and illustrate our insights with simulations of Uber’s market over 137 cities in California.

MB51

Building B Level 3 B311

Online Resource Allocation

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Yue Hu, Stanford University, Stanford, CA, United States

Co-Chair: Shixin Wang, Georgia Institute of Technology, Atlanta, GA, N/A, United States

1 - Survey of Data-Driven Newsvendor: Unified Analysis and Spectrum of Achievable Regrets

Zhuoxin Chen, Tsinghua University, Beijing, China, People's Republic of, Will Ma

In the Newsvendor problem, the goal is to guess the number that will be drawn from some distribution, with asymmetric consequences for guessing too high vs. too low. In the data-driven version, the distribution is unknown, and one must work with samples from the distribution. Data-driven Newsvendor has been studied under many variants: additive vs. multiplicative regret, high probability vs. expectation bounds, and different distribution classes. This paper studies all combinations of these variants, filling in many gaps in the literature and simplifying many proofs. In particular, we provide a unified analysis based on the notion of clustered distributions, which in conjunction with our new lower bounds, shows that the entire spectrum of regrets between $\frac{1}{\sqrt{n}}$ and $\frac{1}{n}$ can be possible. Simulations on commonly-used distributions demonstrate that our notion is the "correct" predictor of empirical regret across varying data sizes.

2 - Online Resource Allocation Without Re-Solving: the Effectiveness of Primal-Dual Policies

Siqi He, Duke University, Durham, NC, United States, Yehua Wei, Jiaming Xu, Sophie Yu

We study an online resource allocation problem where requests arrive sequentially over time and must be served immediately and irrevocably under resource constraints. While state-of-the-art policies, such as probabilistic re-solving (PR) and the Bayes selector, achieve constant regret, they may require re-solving large-scale linear programs (LPs) many times, making them computationally expensive. In this paper, we propose a simple primal-dual policy that requires solving the fluid linear program only once upfront and then uses dynamic dual variable updates to reflect real-time resource availability. Under a general position gap condition equivalent to existing non-degeneracy assumptions, we prove that our policy achieves constant regret independent of the time horizon - the first such guarantee for primal-dual policies in online resource allocation. Our analysis reveals that by adjusting dual variables based on actual resource consumption, the policy maintains a negative drift that pulls the dual variables back toward optimality when they deviate significantly. The policy is computationally efficient, interpretable, and well-suited for distributed implementation as resource constraints are replaced with shadow prices that guide decision-making.

3 - Throughput-Optimal Scheduling Algorithms for LLM Inference and AI Agents

Yueying Li, Cornell University, Ithaca, NY, United States, Jim Dai, Tianyi Peng

As demand for Large Language Models (LLMs) and AI agents rapidly grows, optimizing systems for efficient LLM inference becomes critical. While significant efforts have focused on system-level engineering, little is explored from a mathematical modeling and queuing perspective.

In this paper, we aim to develop the queuing fundamentals for large language model (LLM) inference, bridging the gap between the queuing theory and LLM system communities. In particular, we study the throughput aspect in LLM inference systems. We prove that a large class of 'work-conserving' scheduling algorithms can achieve maximum throughput for individual inference LLM engine, highlighting 'work-conserving' as a key design principle in practice. In a network of LLM agents, work-conserving scheduling alone is insufficient, particularly when facing specific workload structures and multi-class workflows that require more sophisticated scheduling strategies. Evaluations of real-world systems show that Orca and Sarathi-serve are throughput-optimal, reassuring practitioners, while FasterTransformer and vanilla vLLM are not maximally stable and should be used with caution. Our results highlight the substantial benefits that the queuing community can offer in improving LLM inference systems and call for more interdisciplinary development.

4 - Matching Queues, Flexibility and Incentives

Chiwei Yan, University of California Berkeley, San Francisco, CA, United States, Francisco Castro, Peter Frazier, Hongyao Ma, Hamid Nazerzadeh

Agents in online marketplaces (such as ridesharing and freelancing platforms) are often strategic, and heterogeneous in their compatibility with different types of jobs: fully flexible agents can fulfill any job, whereas specialized agents can only complete specific subsets of jobs. Convention wisdom suggests reserving agents that are more flexible whenever possible, however this may incentivize agents to pretend to be more specialized, leading to loss in matches. We focus on designing a practical matching policy that performs well in a strategic environment.

5 - Value of Sparse Structures in Dynamic Reusable Resource Allocation with Waiting

Yue Hu, Stanford University, Palo Alto, CA, United States, Jing Dong, Shixin Wang

We study the dynamic resource allocation problem in online service platforms featuring reusable resources, waiting space, and heterogeneous demands and resources. The service provider aims to balance the trade-off between maximizing revenue and minimizing waiting times across various demand types. For this problem, we propose a comprehensive framework to minimize the long-run average revenue loss and waiting cost by simultaneously designing the 1) flexibility structure, 2) admission control, and 3) scheduling policy. Our proposed flexibility structures are tailored to systems with varying levels of workload intensity, and the number of arcs in the network is linear in the number of resource and demand types. In these sparse networks, we show that simple static priority rules and threshold-based admission controls are asymptotically optimal in the many-server regimes. Furthermore, our proposed algorithm for designing system flexibility, along with the scheduling and admission control policies, is both straightforward to interpret and easy to implement. Numerical experiments demonstrate the effectiveness of our approach, particularly for smaller systems, in non-asymptotic environments.

MB52

Building B Level 3 B312

Centers for Business of Sustainability

Panel Session

Committee's Choice

Co-Chair: Safak Yucel, Georgetown University, Washington, DC, United States

1 - Moderator Panelist

Safak Yucel, Georgetown University, Washington, DC, United States

2 - Panelist

Vishal Agrawal, Georgetown University, Washington, DC, United States

3 - Panelist

Beril Toktay, Georgia Institute of Technology, Atlanta, GA, United States

4 - Panelist

Owen Wu, Kelley School of Business, Indiana University, Bloomington, IN, United States

MB53

Building B Level 3 B313

Information and Algorithm Design in Revenue Management

Invited Session

Revenue Management

Chair: Shouqiang Wang, University of Texas Dallas, Richardson, 75080, United States

1 - Designing Pandora's Box: Product Rankings for Two-Sided Platforms

Shreyas Sekar, University of Toronto - Rotman School of Management, Toronto, ON, Canada, Yiangos Papanastasiou

With the rapid growth of e-commerce, product rankings have emerged as a primary lever used by two-sided platforms to match supply with demand. Although product rankings have been studied from a variety of perspectives, existing research typically focuses on the interaction between ranking policies and consumer search, treating the supply side of the market as exogenous. In this paper, we present a model in which both sides of the market respond strategically to the platform's ranking policy: on the supply side, sellers participate only if their earnings exceed an outside option, while on the demand side, consumers adjust their search behavior to maximize their utility. We illustrate that in such settings, intuitive policies that simply direct consumers to the sellers that best match their preferences are typically suboptimal, resulting in poor marketplace participation. Using tools from the information design literature, we propose a new framework for designing ranking policies that allow the platform to balance supply- and demand-side incentives. Building on this framework, we develop insights into the structure of optimal policies and demonstrate how these insights can be used to design policies which are easy to compute, scalable, and near-optimal in a large set of problem instances.

2 - Stationary Online Contention Resolution Schemes

Mohammad Reza Aminian, The University of Chicago, Booth School of Business, Chicago, IL, United States, Rad Niazadeh, Pranav Nuti

Online resource allocation problems are central to revenue management, where a seller with limited inventory must decide whether to fulfill stochastically arriving requests to maximize expected revenue. A prominent approach to such problems involves solving a fluid relaxation and then applying an *online contention resolution scheme* (OCRS) to implement an online policy. OCRSs not only offer strong performance guarantees but also inherently ensure fairness in allocation—an increasingly critical consideration in many applications.

Despite their appeal, OCRSs are often difficult to design and interpret. We introduce new paradigms that simplify the construction and analysis of OCRSs. First, we define **stationary OCRSs (S-OCRSs)**, which treat each incoming request as if it were the last and maintain a stationary distribution of perceived inventory over time. This framework leads to simple, interpretable allocation policies.

Second, we propose a novel surrogate model where algorithms may serve requests multiple times, which arises naturally from analyzing vanishing probability scenarios. We show how to convert such algorithms into valid OCRSs via a reduction based on *mean-preserving contractions* (MPC).

We apply these frameworks to design near-optimal, greedy-like OCRSs for various settings, including inventory of size k , achieving a selectability guarantee of $1 - 2/\sqrt{2\pi k} + O(1/k)$. Lastly, our MPC framework also provides a new proof of why a greedy algorithm based on the level of available inventory is the optimal policy, which is arguably simpler and more intuitive than the previous proofs using linear programs.

3 - The Design of Intertemporal References with Peer Effects

Lin Chen, INSEAD, Fontainebleau, France, Guillaume Roels

Goals are traditionally classified into two types: those based on task-specific or *intrapersonal* standards of competence, and those based on *interpersonal* standards of competence. The former relies on one's own past performance, while the latter is strongly shaped by the availability of information about others. Using a dynamic framework, we characterize how individual effort responds to intra- and interpersonal goals. Building on this, we demonstrate how performance transparency can serve as an intervention to promptly enhance agents' performance.

4 - Recommender Systems Under Privacy Protection

Shouqiang Wang, University of Texas Dallas, Richardson, TX, United States, Can Kucukgul, Ozalp Ozer

We study how a profit-driven online platform designs its recommender policies in response to different privacy protection regimes that grant users varying degrees of control over their personal data. We demonstrate the effective equivalence between the opt-out protection and the unprotected privacy. As a key finding, consumers' autonomy over their privacy (to the extent that they could misrepresent their personal data) may compel platforms to distort their recommender policies and lead to unintended consequences. When the recommendation plays only an informative role, such level of privacy protection deters the platform from any personalized recommendation. If the recommendation can additionally play an allocative role to control consumers' access to products, the interplay between the recommendation's informative and allocative roles may give rise to algorithmic discrimination, whereby the disadvantaged minority in the society are restricted or deprived of access to potential valuable opportunities. Ultimately, these distortions could inadvertently hurt both platforms and consumers, relative to less stringent privacy protection regimes. Counter-intuitively, enacting the recommendation's allocative role that could limit users' access to certain products in addition to its informative role can in fact benefit both the platform as well as the users, especially when users are given the autonomy over their privacy. Our findings are robust even when the products' profit margins are endogenous.

MB54

Building B Level 2 B216

Empirical Approaches to Retail Decision-Making

Invited Session

Revenue Management

Chair: Nymisha Bandi, Stony Brook University, Stony Brook, NY, H3V1G2, United States

1 - Restaurant Density and Delivery Speed in Food Delivery Platforms

Wenchang Zhang, Indiana University Bloomington, Bloomington, IN, United States, Zhanzhi Zheng, Ruomeng Cui

With the rapid growth of food delivery platforms, understanding their operational implications for partnered restaurants is crucial. This paper compares three main externalities induced by an increase in restaurant density on a platform: customer expansion, cannibalization, and delivery pooling (i.e., drivers being in closer proximity to restaurants). Using a theoretical model, we develop testable hypotheses about the impact of these externalities on various aspects of partnered restaurants, including delivery speed (i.e., order wait time), sales, revenues, and customers' spending behaviors. Econometric analysis, based on data from a major Chinese food delivery platform, shows that as the platform's restaurant density increases, delivery speed, sales, and revenues improve for partnered restaurants. However, customers are found to opt for smaller order sizes in denser areas. The empirical evidence then indicates that delivery pooling emerges as the dominant externality compared to customer expansion and cannibalization as a platform expands regionally. We further estimate that the driver arrival speed at a restaurant and restaurant density follow an inverted-U-shaped relationship, suggesting a diminishing benefit to delivery speed from increasing restaurant density. Lastly, using instrumental variable estimation, we find that both restaurant sales and customer spending behaviors are highly sensitive to delivery speed, underscoring the dominant role of delivery pooling. Our findings highlight the importance of optimizing restaurant density for food delivery platforms undergoing expansion and recommend a focus on maintaining high restaurant density areas to maximize delivery speed. This can foster a scenario beneficial not only for platforms but also for restaurant partners and customers.

2 - Voice Chatbot Design: Leveraging the Preemptive Prediction Algorithm

Shuai Hao, Southern University of Science and Technology, Shenzhen, IL, United States, Yuqian Xu

This paper investigates the design of artificial intelligence (AI) voice chatbot systems by introducing a novel preemptive prediction feature that enhances human-AI interaction through workflow innovation. Partnering with a leading global e-commerce and logistics platform, we examine the design and performance of a novel preemptive prediction feature that proactively identifies the likely package for a customer's inbound inquiry call at the outset of the AI chatbot interaction. Interestingly, although the prediction feature does not modify the underlying technical logic used by the chatbot to resolve service inquiries, it meaningfully influences service outcomes through behavioral channels. Using detailed call-level data and a difference-in-differences estimation framework, we find that the feature improves both productivity by reducing average call duration by 5.3% and service quality by raising inquiry resolution rates by 5.0% and customer satisfaction by 4.6%. However, when predictions are incorrect, service performance deteriorates, though the losses are smaller than the gains from correct predictions. Further analysis reveals that customers with more prior

exposure to the AI chatbot benefit more from the feature, while rural users and those expressing strong initial sentiment (either positive or negative) are more adversely affected by incorrect predictions. To uncover the behavioral mechanisms underlying this effect, we show that correct predictions significantly reduce the likelihood of transfers to human agents, thereby increasing the probability that the chatbot has an opportunity to complete the service request.

3 - Bundling OR Discounting? Field Experiments for Healthy and Unhealthy Food

Nymisha Bandi, McGill University, Montreal, QC, Canada

Retailers routinely face the strategic decision of how to promote their products. Two of the most common promotion strategies are bundling and discounting. Determining which strategy is more effective, and how their effectiveness varies across product categories, is an open question in the world of retail. To address this, we conducted a multi-store field experiment with a global convenience store chain that regularly uses bundle promotions (particularly for certain unhealthy snacks). Our study focused on healthy and unhealthy food categories to understand how these promotions influence consumer behavior and financial outcomes. Specifically, we tested these strategies---discounts and bundles---applied to (i) healthy snacks, (ii) unhealthy snacks, and (iii) both categories simultaneously. For healthy snacks, we found that while discounting was more effective at boosting sales, bundling was more effective at increasing revenue and profit. Additionally, both promotion strategies for healthy snacks increased the sales of other unhealthy items, as consumers may engage in moral licensing, using their healthy purchases as justification to indulge in less healthy options. In contrast, neither promotional strategy had an effect on sales for unhealthy snacks. Based on this surprising finding, we convinced the retailer to conduct a follow-up experiment to investigate the impact of discontinuing a long-standing bundle promotion on unhealthy snacks. The results of this second experiment revealed no change in sales but an increase in revenue and profit, confirming our initial conclusion. These insights provide valuable guidance for retailers seeking to optimize promotion strategies for different product categories.

4 - Estimating Demand in the Absence of Sales and Inventory Information

Sarang Deo, Indian School of Business, Hyderabad, India, Sripad Devalkar, Aditya Jain

Unorganized retail stores in emerging economies, unlike conventional organized retail stores, are rarely equipped with point-of-sale devices. This leads to lack of data on retail sales and inventory positions that are crucial for effective upstream decisions made by the supplier. Existing studies in the area of inventory management have assumed availability of censored demand data (in the form of sales observations) and inventory data (order upto levels, inventory positions, stockout indicators, etc.). Our study, being pertinent to the context of unorganized retail, assumes the availability of only replenishment data. We begin by observing that replenishment is related to downstream sales through the retailer's inventory policy. We make structural and parametric assumptions about demand and inventory policy, and develop a methodology built on the Expectation-Maximization algorithm to jointly estimate demand and the inventory policy of the retailer.

MB55

Building B Level 3 B314

Advances in Choice-Based Revenue Management and Assortment Optimization

Invited Session

Revenue Management

Chair: Antoine Desir, INSEAD, PARIS, France

1 - Assortment Optimization with Downward Feasibility: Efficient Heuristics Based on Independent Demands

Guillermo Gallego, Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Pin GAO, Shaoyu Wang, Gerardo Berbeglia

A core challenge in revenue management is to select a subset of products or services to offer customers to maximize a firm's expected revenue. Given the complexity of optimizing assortment decisions under intricate consumer behavior and business constraints, industry practitioners often resort to estimating product demand and optimizing decisions based on the independent demand (IDM) assumption, namely each product's demand is independent of the other offered products. To address its limitation of overlooking substitution effects, this paper proposes an IDM-based optimization variant that aligns "independent" demands with purchase probabilities for single-product offerings and uniformly modifies product revenues by a constant. For last-choice regular models that meet certain bounding conditions, our heuristic achieves the best possible performance guarantees and necessitates solving only two linear programs when the constraints are totally unimodular (TUM) and downward feasible. More generally, we propose heuristic extensions for cases where constraints are not TUM or the bounding conditions are difficult to determine. Beyond its practical simplicity, the proposed methodology also facilitates the development of efficient heuristics for several previously studied problems, either by relaxing assumptions or improving approximation ratios. Extensive computational experiments demonstrate that the proposed heuristics consistently deliver exceptional performance. Furthermore, analysis of a real-world dataset reveals that our heuristics, when derived from models with higher predictive accuracy, achieve superior optimization outcomes.

2 - Two Sides of "Hidden City Ticketing": Analysis of a Choice-Based Network Revenue Management Model

Nanxi Zhang, Ivey Business School, University of Western Ontario, London, ON, Canada, Jiacheng Guo, Bo Jiang, Zikai Xiong

"Hidden City Ticketing" (HCT) is a strategy where customers book flights with layovers at their intended destinations and skip the final leg of the journey. Empirical studies on airfare data have shown that HCT chances are very prevalent (about 15%-25%) in airline pricing. In this paper, we investigate the impact of HCT on airline revenue and discuss if the airline should eliminate all HCT chances from its price to achieve higher profit. We use a choice-based network revenue management model to characterize the airline's dynamic pricing problem. As airlines may penalize customers who purchase HCT by canceling frequent flyer miles and banning future travel with the airline, we capture customers' concerns about the risks of HCT through a choice model. Our analysis shows that adjusting prices to completely eliminate HCT chances may overly constrain the pricing strategy, and it harms the airline revenue more than customers' HCT purchase behavior. Furthermore, the impact of HCT is two-sided: While HCT can erode the airline's revenue, it can also boost revenue in some cases. In the worst case, if customers purchase HCT, the airline's revenue can be half of what it would be if customers don't purchase HCT. In the best case, the airline's revenue in the case customers purchase HCT can be twice that of the case where customers do not purchase HCT. Furthermore, the worst-case and best-case revenue bounds are both tight.

3 - Constrained Two-Sided Assortment Optimization

Ulysse Hennebelle, Cornell University, Ithaca, NY, United States, Omar El Housni, Alfredo Torrico

Two-sided matching platforms have proliferated thanks to their application in labor markets, dating, accommodation and ride-sharing. These platforms face the challenge of reducing choice congestion among popular options who receive more requests than they can handle, which leads to suboptimal market outcomes. To address this challenge we introduce a two-sided assortment optimization framework under general choice preferences and constraints over the assortments. The goal in this problem is to maximize the expected number of matches by deciding which assortments are displayed to the agents and the order in which they are shown. In this context, we identify several classes of policies that model the different levels of adaptivity that platforms can use in their design. Our goals are: (1) to measure the value that one class of policies has over another one, and (2) to solve the optimization problem itself for a given class.

For (1), we characterize the gaps between all pairs of classes, under any choice model and various types of constraints. These results showcase the benefits of each class of policies and show that a vast group of them ranges within a constant factor of the most adaptive optimal policy.

For (2), we provide algorithms that achieve constant approximation for every policy class, both in the unconstrained and constrained settings. In addition to studying various policy classes, this work is also the first to address constrained two-sided assortment optimization where the constraints apply to both sides of the market, which unveils new significant algorithmic challenges.

4 - Learning-Based Pricing Policy for Substitute Itineraries

Yue Su, Université de Lille, Inria, CRISTAL, Villeneuve-d'Ascq, France, Antoine Desir, Axel Parmentier

In airline revenue management, dynamic pricing for substitutable itineraries presents significant challenges due to the stochastic nature of customer choices. The dynamic pricing problem can be conceptualized as a Markov Decision Process with the objective of maximizing the total expected revenue over a finite selling horizon. However, the complexity of this model, characterized by its multi-dimensional state and action spaces, makes it computationally prohibitive to solve exactly, even for small-to-medium-sized instances. As a result, classic methods such as dynamic programming (DP) prove inefficient in addressing the dynamic pricing problem for larger instances.

In this work, we introduce learning-based policies tailored to the dynamic pricing problem of substitute itineraries. We proposed two different machine learning (ML) pipelines to encode pricing policy and designed tailored training processes for each of them. Compared to existing methods, our approach has two advantages: (1) benchmarking the proposed policies against the optimal pricing policy, we show that learning-based policies seem to be a good approximation for the optimal policy; (2) our learning-based policies can lead to huge speed-up in solving large-scale instances (with up to million-magnitude state space) by moving all CPU time offline.

MB56

Building B Level 3 B315

Machine Learning and Human Incentives

Invited Session

Auctions and Market Design

Chair: Chara Podimata, MIT, Cambridge, MA, United States

Co-Chair: Ellen Vitercik, Stanford University, Stanford, CA, United States

Co-Chair: Nicolas Emmenegger

1 - Learning-Augmented Mechanism Design

Xizhi Tan, Drexel University, Philadelphia, PA, United States

In the framework of "algorithms with predictions," algorithms are augmented with a machine-learned prediction and the goal is to obtain improved guarantees when the prediction is correct (consistency) while simultaneously guaranteeing some worst-case bounds even if the prediction is arbitrarily wrong (robustness). The majority of the work on this framework has focused on online algorithms with predictions regarding future input. A subsequent line of work has focused on mechanism design, where the prediction is regarding the private information of strategic agents. In this talk, we initiate the study of online mechanism design with predictions, which combines the challenges of online algorithms with predictions and mechanism design with predictions. We consider the problem of designing a revenue-maximizing auction to sell a single item to strategic bidders who arrive and depart over time. We study the learning-augmented version of this problem, where the auction designer is given a prediction regarding the maximum value over all agents. Our main result is a strategyproof mechanism whose revenue guarantees are α -consistent with respect to the highest value and $(1-\alpha^2)/4$ -robust with respect to the second-highest value, for $\alpha \in [0,1]$. We show that this trade-off is optimal within a broad family of auctions.

2 - Gap Based Learning in Markets

Rigel Galgana, Massachusetts Institute of Technology, Cambridge, MA, United States

We consider a broad class of learning algorithms---which we dub "gap based algorithms" (GBAs)---which parameterize a learner's sensitivity and tolerance to actions' sub-optimality. At a high level, GBAs guarantee that the actions with large sub-optimality are selected with vanishing probability (pruned) while those with small sub-optimality are explored frequently. As these natural conditions are satisfied by many popular learning algorithms such as Hedge and epsilon-greedy, GBAs are ubiquitous in both theory and practice. In this paper, we analyze the game dynamics induced by GBAs in several important economic models of competition include: 1) the first price auction, 2) Bertrand pricing competition, and 3) the multi-unit uniform price auction. With some additional conditions on the GBAs' exploration and pruning rates, we show that the corresponding dynamics necessarily escape unstable equilibria. Our results have interesting implications for equilibrium selection and evolutionary stability.

3 - Learning to Price Electricity for Optimal Demand Response

Jing Shang, Stanford University, Stanford, CA, United States, Xinyang Zhou, Mohammad Mehrabi, Yue Chen, Moody Saleh, Andrey Bernstein, Stefan Wager

There is considerable interest in using time-varying electricity prices to drive consumer demand response, and better align energy demand with renewable production. However, optimal prices generally vary over time in response to complex signals such as weather forecasts, sunrise/sunset times, and day-of-week patterns; and existing methods are not able to make efficient use of such rich contextual information. Here, we propose a neural-network-based algorithm for contextual energy pricing. Our approach models the pricing problem as a Stackelberg game, and leverages a representation of its mean-field solution given in Mehrabi et al.(2024) for learning. We validate our approach by simulating the energy grid in several cities, and show that incorporating contextual information for pricing can considerably increase the value of the demand response programs.

4 - Strategy-Aware Regression Discontinuity Designs

Nicolas Emmenegger, MIT, Cambridge, MA, United States, Chara Podimata

In this work, we review regression discontinuity designs (RDDs) within a non-asymptotic minimax statistical framework using the tools of incentive-aware machine learning. We give a precise characterisation of the information complexity of estimating local treatment effects under a Lipschitz smoothness assumption. We do so under four agent behavioral models, and show that each model corresponds to the best response of a principal-agent Stackelberg game. Interestingly, we show that the minimax complexity depends on the behavioral model itself. Some models lead to the same information complexity as the non-strategic case, while some lead to irreducible bias -- and some fall in between leading to slower rates. We provide computationally tractable estimators for each setting, and demonstrate their robustness to misspecification using synthetic data. Finally, we apply our results to a popular dataset from economics, suggesting that strategic agents can skew the results of empirical studies -- and that our methods can help alleviate this.

MB57

Building B Level 3 B316

From Digital Signals to Strategic Decisions: AI, Social Media, and the Future of Business Analytics

Invited Session

Social Media Analytics

Chair: Cuibing Wu, Framingham State University, 100 State St, Sherborn, MA, 01701, United States

Co-Chair: Esmat (Yasi) Sangari, University of San Francisco, School of Management, San Francisco, CA, United States

1 - Bridging the Gap: Advancing Diversity, Equity, and Inclusion in Corporate Innovation Through AI and Big Data Analysis

Cuibing Wu, Framingham State University, Framingham, MA, United States

This study analyzes how corporate earnings calls shape diversity, equity, and inclusion (DEI) outcomes using a dataset of over 170,000 transcripts (2013–2023) from SeekingAlpha. Leveraging large language models (LLMs) trained on DEI and innovation lexicons, we extract and assess DEI discourse through techniques such as keyword extraction, topic modeling, and sentiment analysis. We focus on four DEI dimensions: demographic representation, inclusivity initiatives, equity efforts, and future DEI outlooks. Regression models link these to innovation outcomes, revealing which communication patterns predict stronger DEI and innovation performance. The findings offer strategic insights for promoting DEI through corporate communications.

2 - Tracing Early Signals of Educational Demand Shifts in the Era of Generative AI

Ruilin Wang, UMass Lowell, Lowell, MA, United States, Ben Li

The emergence of generative artificial intelligence (AI), particularly large language models, is fundamentally altering public perceptions of the value and relevance of traditional educational pathways. As these technologies increasingly undertake complex cognitive tasks once reserved for highly educated professionals, they prompt pressing questions about which academic disciplines will remain resilient and which may face obsolescence. While prior research has largely centered on labor market disruptions or lagging institutional metrics such as enrollment trends, our study redirects attention to early-stage behavioral responses following a major AI breakthrough. By analyzing longitudinal shifts in public interest across a broad spectrum of academic programs and timeframes, we offer a novel perspective on how technological shocks shape educational demand in real time—informing institutional strategy, curriculum development, and policy adaptation.

3 - Open-Source AI Safe Engagement: An Approach to Discovering Highly Adopted and Capable Open-Source Models

Esmat Sangari, University of San Francisco, School of Management, San Francisco, CA, United States

As the number of open-source AI models on platforms like Hugging Face continues to grow, identifying which models will gain widespread adoption and ensuring they meet safety standards has become a critical challenge. In this study, we examine the dynamics of model adoption in the open-source AI ecosystem and propose a framework for predicting model popularity using a set of interpretable features. Our approach provides a scalable way to prioritize evaluation efforts and safety audits on high-impact models. The findings offer practical insights for developers, users, and platform designers aiming to promote responsible and effective participation in open-source AI communities.

4 - A Theory-Based Explainable Graph Attention Network for Detecting Misleading Forward-Looking Statements in Earnings Conference Calls

Elaheh Aghazadeh, UMass Lowell, Lowell, MA, United States

Earnings conference calls are a key venue through which corporate executives communicate forward-looking statements to investors. While these statements can offer valuable insights into a firm's future prospects, they are also prone to strategic obfuscation and linguistic manipulation that may mislead stakeholders. This study develops a theory-based, explainable Graph Attention Network (GAT) model to detect potentially misleading forward-looking statements in earnings calls and assess their relationship with subsequent financial fraud. Integrating insights from signaling theory, impression management, and deception detection, the proposed model leverages theory-informed linguistic features, such as hedging, over-positivity, and vagueness, alongside pre-trained embeddings. The dialogue structure of earnings calls is modeled as a heterogeneous graph consisting of speaker and utterance nodes, with edges encoding conversational and role-based relationships. Attention mechanisms within the GAT architecture provide interpretable insights into which linguistic cues and speaker interactions drive model predictions. Using a weak supervision strategy informed by SEC enforcement actions (AAERs), we construct a case-control dataset and evaluate the model's ability to (1) detect misleading statements and (2) predict future fraud at the firm level. The findings aim to enhance both detection accuracy and transparency, offering practical value to regulators, analysts, and investors seeking to identify early warning signs of corporate misconduct.

MB58

Building B Level 3 B317

Manufacturing Analytics

Contributed Session

Contributed

Chair: Ranli Lu, 1020 Luxe Ave Apt8408, 43220

1 - Analytical Prediction of Optimized Processing Parameters in Metal Additive Manufacturing

Wei Huang, Georgia Institute of Technology, Atlanta, GA, United States

Metal additive manufacturing (AM) has emerged as a transformative technology for producing complex geometries with tailored mechanical properties. However, achieving optimal part quality and performance requires precise control over processing parameters such as laser power, scan speed, layer thickness, and hatch spacing. This study presents an analytical framework for predicting optimized processing parameters in metal AM, focusing on the interplay between thermal dynamics, microstructure evolution, and mechanical properties. The proposed approach integrates thermal modeling, solidification kinetics, and machine learning algorithms to identify optimized parameter sets that maximize mechanical strength. The framework is validated through experimental data from laser powder bed fusion (LPBF) of titanium alloys, demonstrating its capability to predict parameter combinations that yield superior part quality. This work provides a systematic pathway for accelerating the development of optimized metal AM processes, reducing trial-and-error experimentation, and enhancing the reliability of AM-produced components.

2 - Congestion-Focused Approach to Semiconductor Wafer Fabrication Facility Layout

Gwangjae Yu, Southern Illinois University Edwardsville, Edwardsville, IL, United States, Hyunsuk Baek, John Fowler

In a facility layout design problem for semiconductor fabrication facilities, the traditional approach has been to minimize the flow-weighted distance of material flow. Transports are conducted by an Automated Material Handling System on its internal transportation network. However, this approach often leads to significant traffic congestion, raising the question of whether minimizing flow-weighted distance is truly an effective criterion for designing facility layouts. This study proposes a novel approach to layout design for semiconductor fabrication facilities, termed "flow-balancing", to mitigate traffic congestion within the transportation system by leveraging the concept of entropy to quantify congestion in the transportation network. We present Mixed Integer Linear Programming models for designing the facility layout of semiconductor fabrication facilities under both the traditional and proposed approaches, and compare their performances using discrete event simulation. The result shows that the flow-balancing approach tends to outperform the traditional method in key performance measures of transportation system in semiconductor manufacturing facilities, with improvements of up to 11% in the total number of deliveries.

3 - A Three Dimensional Pipe Routing Problem in Semiconductor Fabrication Industry

Junghwa Lee, Korea University, Seoul, Korea, Republic of, Wonhee Cho, Hyungjoo Cha, Taesu Cheong

In this paper, we propose a three-dimensional shortest path problem involving a single origin and multiple candidate destinations for exhaust pipe routing in the semiconductor fabrication industry. While previous studies have actively addressed variations of the shortest path problem, this study focuses on extending the problem into three-dimensional space by incorporating calculations for pressure loss due to pipe length and multiple types of bend angles, obstacle avoidance, and safety standards relevant to industrial applications. To address the proposed problem, we develop a mixed-integer linear program. Noticing the problem complexity, we propose Pressure-based A* algorithm with Divide-and-Conquer strategy (PA*-DC) that utilizes 3D Pipe Routing A* (3DP-A*) algorithm iteratively. We validate the algorithms through

computational experiments using a real-world dataset provided by Samsung Electronics and compare the results with current industry de facto standards for pipe design in actual fabrication facilities.

4 - Economic Evaluation of Jit Systems

Ranli Lu, The Ohio State University, Columbus, OH, United States, Marc Posner

Just-in-Time (JIT) and Material Requirements Planning (MRP) are widely used production systems, each with specific strengths and weaknesses. While JIT favors lean practices and continuous improvement, it has difficulties when demand varies widely. In contrast, MRP is more effective when there is considerable demand variability. We explore how these two systems perform across a range of production scenarios. We derive cost formulas and analytical bounds that allow for direct comparison between the two systems. This provides practical guidance for manufacturers when deciding whether it may be economic to replace an existing JIT system with an MRP system.

MB60

Building B Level 3 B319

Applications of AI and Analytics

Contributed Session

Contributed

Chair: Chengyuan Deng, Rutgers University, New Brunswick, NJ, United States

1 - Optimizing Patient Logistical Decisions: Addressing the Overlooked Problem of Digital Blindness and Infrastructure Gaps in Healthcare Information Systems

Ipe Paramel, Oklahoma State University - Center for Health Systems Innovation, Tulsa, OK, United States, Marjorie Erdmann, William Paiva

Across healthcare systems, patient logistics are managed via technology platforms. In a mixed methods quality improvement (QI) study, working to reduce emergency department (ED) boarding, we identified a consistent pattern of suboptimal logistical decisions across departments, stemming not from human error, but from incomplete or poorly adapted information in electronic medical record (EMR) systems and related tools. We define this systemic issue as digital blindness (DB). By comparing delays caused by DB to other root causes, we found DB to be the leading contributor to boarding-related inefficiencies. Following targeted interventions to resolve DB, we compared pre- and post-intervention periods using logistic regression, controlling for ED volume and bed occupancy. Results showed a 13-minute per patient reduction in boarding time. Run chart analysis confirmed sustained improvements over seven consecutive months, increasing ED capacity by 689 hours over a nine-month period. This study underscores the importance of designing health information systems that align with real-time decision-making needs of frontline staff. We are now identifying DB types, mapping where and how DB erodes care delivery and hospital throughput, and testing efficient methods to resolve DB. Looking forward, we propose leveraging AI tools to monitor, detect, and preempt DB in real time, enabling proactive infrastructure improvements before patient care is compromised. This work aims to lay the foundation for intelligent, adaptive systems that optimize patient logistics across complex healthcare environments.

2 - A Data-Driven County-Level Budget Allocation Model to Manage the Opioid Crisis in West Virginia

Mahdi Saadati, Binghamton University, Binghamton, NY, United States, Ankit Bansal, Jean-Philippe Richard, Sam Workman, Gordon Smith, Henry Brownstein

West Virginia (WV) has one of the highest per-capita opioid-related mortality rates in the United States, reaching 80.9 deaths per 100,000 people in 2022. This study presents a data-driven optimization framework aimed at guiding county-level budget expenditures in WV to reduce opioid-related deaths (ORDs) and crime incidents (CIs). We begin by assembling a panel dataset covering county-level expenditures, ORDs and CIs from 2012 to 2022. To identify causal relationships between the county-level budget expenditures, ORDs, and CIs, we apply the Arellano-Bond Dynamic Panel Model. Next, we train a Tree Ensemble (TE) model to predict ORDs, and CIs based on budget expenditures and various socio-economic indicators and embed it into a Mixed Integer Linear Programming (MILP) framework, creating a robust utility maximization model to determine optimal budget expenditures. Focusing on 12 southern counties—WV's region most severely impacted by the opioid epidemic—our model suggests that different budget allocations could potentially reduce both ORDs and CIs significantly.

3 - Fedgem: a Federated Generalized Expectation-Maximization Algorithm for Mixture Models with An Unknown Number of Components

Michael Ibrahim, Georgia Institute of Technology, Atlanta, GA, United States, Nagi Gebraeel, Weijun Xie

We study the problem of federated clustering when the total number of clusters across clients is unknown, and the clients have heterogeneous but potentially overlapping cluster sets in their local data. To that end, we develop FedGEM: a federated generalized expectation-maximization algorithm for the training of mixture models with an unknown number of components. Our proposed algorithm relies on each of the clients performing EM steps locally, and constructing an uncertainty set around the maximizer associated with each local component. The central server utilizes the uncertainty sets to learn potential cluster overlaps between clients, and infer the global number of clusters via closed-form computations. We present a thorough theoretical study of our algorithm, presenting probabilistic convergence guarantees under common assumptions. Subsequently, we study the specific setting of isotropic GMMs, providing tractable, low-complexity computations to be performed by each client during each iteration of the algorithm, as well as verifying assumptions required for algorithm convergence. We perform various numerical experiments, where we empirically demonstrate that our proposed method achieves comparable performance to centralized EM, and that it outperforms various existing federated clustering methods.

4 - Optimizing Public Healthcare Investments: a Clustering-Based Approach to Reducing Waiting List TIMES in Chile

Constanza Contreras, Universidad de Chile, Santiago, Chile, Patricio Wolff, Sebastián Ríos Pérez

In Chile, the challenge of managing waiting lists in the public healthcare system has been approached through multiple strategies, including the injection of public funds to procure services from the private sector. However, the effectiveness of these interventions in reducing waiting times remains variable and often suboptimal. In this study, we present a novel optimization model designed to strategically identify clusters of patients, pathologies, and locations that yield the greatest reduction in median waiting time per unit of investment. We collected the discharges of elective surgeries in the Chilean waiting list from 2018 to 2024, which contains patient demographics, diagnosis, and intervention details, as well as the historical DRG codification. Central to our approach is the development of a cost estimation methodology that quantifies the expected expenditure associated with the resolution of each individual case on the waiting list. This enables more informed and targeted allocation of financial resources.

Our model was validated through a comparative analysis with historical budget allocation strategies. We compared the results according to expected health outcomes and maximal private sector capacity. This work contributes to the field of healthcare operations research by offering a data-driven tool for optimizing service delivery and enhancing the impact of public health investments.

5 - Nonparametric Classification on Low Dimensional Manifolds Using Overparameterized Convolutional Residual Networks

Zixuan Zhang, Georgia Tech, Atlanta, GA, United States

Convolutional residual neural networks (ConvResNets), though overparameterized, can achieve remarkable prediction performance in practice, which cannot be well explained by conventional wisdom. To bridge this gap, we study the performance of ConvResNeXts trained with weight decay, which cover ConvResNets as a special case, from the perspective of nonparametric classification. Our analysis allows for infinitely many building blocks in ConvResNeXts, and shows that weight decay implicitly enforces sparsity on these blocks. Specifically, we consider a smooth target function supported on a low-dimensional manifold, then prove that ConvResNeXts can adapt to the function smoothness and low-dimensional structures and efficiently learn the function without suffering from the curse of dimensionality. Our findings partially justify the advantage of overparameterized ConvResNeXts over conventional machine learning models.

6 - AI-Driven Dynamic Scenario Planning for Demand Forecasting

Chongwoo Park, Korea University Business School, Seoul, Korea, Republic of, Hosun Rhim

Accurate demand forecasting is critical for effective operations management, particularly in environments characterized by rapid innovation and volatile markets. Conventional forecasts extrapolate historical averages and linear trends, overlooking the nonlinear impacts of disruptive technologies, policy shifts, and evolving consumer preferences. This study proposes an AI-driven Dynamic Scenario Planning (DSP) architecture that embeds Bass Diffusion and Market Share models within the VISFER framework. The procedure proceeds through five stages: (1) influential variables—consumer taste, product attributes, cost structures, and macro indicators—are isolated through machine-learning screens; (2) system-thinking tools translate these drivers into causal networks that anchor scenario construction; (3) artificial intelligence generates diverse, data-rich narratives spanning technological, regulatory, and competitive uncertainty; (4) quantitative forecasting combines Bass diffusion, capturing innovators and imitators, with a market-share module that traces competitive redistribution during maturity; and (5) iterative evaluation updates parameters as empirical feedback accrues. Comparative experiments demonstrate that the integrated DSP approach reduces mean absolute percentage error and forecast bias relative to single-scenario baselines, while supplying granular insights into adoption pathways. Consequently, managers obtain actionable guidance for capacity sizing, resource allocation, and risk buffering under uncertainty. By enabling proactive identification of demand inflections, the architecture optimizes supply-chain operations and mitigates costs associated with over- or under-production. Beyond its practical benefit, the study contributes to operations-management scholarship by delivering a scalable, continuously learning forecasting platform calibrated for dynamic markets. Overall, the findings highlight the strategic value of fusing AI-driven analytics with scenario planning to support decisions in complex, uncertain environments across industries confronting relentless global change.

7 - On the Price of Differential Privacy for Hierarchical Clustering

Chengyuan Deng, Rutgers University, New Brunswick, NJ, United States

Hierarchical clustering is a fundamental unsupervised machine learning task of organizing data into a hierarchy of clusters. Many applications of hierarchical clustering involve sensitive user information, therefore motivating recent studies on differentially private hierarchical clustering under the rigorous framework of Dasgupta's objective. However, it has been shown that any privacy-preserving algorithm under edge-level differential privacy necessarily suffers a large error. To capture practical applications of this problem, we focus on the weight privacy model, where each edge of the input graph is at least unit weight. We present a novel algorithm in the weight privacy model that shows significantly better approximation than known impossibility results in the edge-level DP setting. In particular, our algorithm achieves $O(\log^{1.5} n / \epsilon)$ multiplicative error for ϵ -DP and runs in polynomial time, where n is the size of the input graph, and the cost is never worse than the optimal additive error in existing work. We complement our algorithm by showing if the unit-weight constraint does not apply, the lower bound for weight-level DP hierarchical clustering is essentially the same as the edge-level DP, i.e. $\Omega(n^2 / \epsilon)$ additive error. As a result, we also obtain a new lower bound of $\Omega(1 / \epsilon)$ additive error for balanced sparsest cuts in the weight-level DP model, which may be of independent interest. Finally, we evaluate our algorithm on synthetic and real-world datasets. Our experimental results show that our algorithm performs well on extra cost and has good scalability to large graphs.

MB61

Building B Level 4 B401

Spatio-temporal Modeling and Forecasting for Energy Systems

Invited Session

Quality, Statistics and Reliability

Chair: Ahmed Aziz Ezzat, Rutgers University, Piscataway, NJ, United States

1 - Gaussian Processes (GPS) for Spatial-Temporal Power Curve Estimation

Ahmadreza Chokhachian, ISyE Georgia Tech, Atlanta, GA, United States, Roshan Joseph, Yu Ding

In this study, we address the challenge of modeling wind power curves by incorporating the effects of terrain. Unlike environmental variables such as wind speed and temperature, which vary both spatially and temporally, terrain features such as ruggedness, slope, and ridges vary between turbine locations but remain constant over time. To effectively handle these two modes of variation, we employ a Kronecker product Gaussian Process (GP) model. By jointly modeling data from multiple turbines, this approach integrates terrain information into the power curve estimation process, leading to more accurate, site-specific predictions. Moreover, it enhances the model's generalizability to turbines located in different terrains, underscoring the value of incorporating spatially-varying terrain data in wind energy forecasting.

2 - Online Changepoint Detection for Circular Data with Applications to Wind Direction

Tianyi Xu, University of Michigan, Ann Arbor, MI, United States

Online changepoint detection in circular data, such as wind direction, plays an important role in energy systems applications, including wind farm operations and wildfire risk monitoring. However, most existing online detection methods are tailored to linear data and are not directly applicable to circular domains. While a few recent studies address circular data, they often focus on detecting a single changepoint under prespecified in-control and out-of-control states. In this work, we extend an existing Bayesian online changepoint detection framework to accommodate sequential circular data. By integrating the von Mises distribution, we derive sequential updating rules and develop an approximation scheme to enable efficient online detection of multiple changepoints. Numerical experiments and a case study using wind direction data demonstrate that our method achieves low false alarm and missed detection rates while maintaining detection accuracy and responsiveness compared to existing benchmarks. This approach supports more responsive monitoring of spatio-temporal patterns in energy systems.

3 - Spatio-Temporal Forecasting for the Next-Generation of Wind Turbines: a Deep Statistical Method

Ahmed Aziz Ezzat, Rutgers University, Piscataway, NJ, United States, Feng Ye, Xinxi Zhang, Michael Stein

To unlock access to stronger winds, the offshore wind industry is advancing with significantly larger and taller wind turbines. Existing spatio-temporal forecasting methods compress the three-dimensional reality of the wind field (space $\times$ time $\times$ height) into a simpler, two-dimensional “world view” (space $\times$ time) which completely overlooks the vertical dimension of the wind field. Instead, we propose a multi-output spatio-temporal model with a multivariate, nonstationary, and state-dependent kernel that is capable of adequately modeling the complex dependencies in local wind speeds across multiple heights. Embedded within the statistical model, an advanced deep learning architecture fully exploits the wealth of exogenous information available to the forecaster (typically in the form of high-dimensional image streams) in order to learn and update the kernel parameters. Experiments using real-world data from the offshore wind energy areas in the Northeastern United States show that the wind speed and power forecasts from the proposed model are of considerably higher quality than those from prevalent time series, deep learning, and space-time methods.

4 - Enhanced Renewable Energy Forecasting Using Physics-Informed Conformal Prediction

Alireza Moradi, Georgia Tech, Atlanta, GA, United States, Mathieu Tanneau, Reza Zandeshahvar, Pascal Van Hentenryck

Accurate forecasting is critical for reliable power grid operations, particularly as the share of renewable generation—such as wind and solar—continues to grow. Given the inherent uncertainty and variability in renewable generation, probabilistic forecasts have become essential for informed operational decisions. However, such forecasts frequently suffer from calibration issues, potentially degrading decision-making performance. Building on recent advances in Conformal Predictions, this paper introduces a tailored calibration framework that constructs context-aware calibration sets using a novel weighting scheme. The proposed framework improves the quality of probabilistic forecasts at the site, regional and fleet levels, as demonstrated by numerical experiments on large-scale datasets covering several systems in the United States. The results demonstrate that the proposed approach achieves higher forecast reliability and robustness for renewable energy applications compared to existing baselines.

5 - Spatio-Temporal Statistical Analysis of Power Grid Resilience to Extreme Weather Conditions

Apoorva Narula, Georgia Institute of Technology, Atlanta, GA, United States, Feng Qiu, Yao Xie

We present a spatio-temporal statistical modeling framework to assess power grid resilience under extreme weather events, with a focus on hurricanes. Leveraging large-scale outage and meteorological data, we quantify how outage events propagate and recover across geographic regions. Our case study focuses on the Eastern and Southeastern U.S. states most affected by hurricanes between 2014 and 2022: North Carolina, South Carolina, Florida, Alabama, Mississippi, Louisiana, and Texas.

Outage events are defined as contiguous time periods with non-zero reported customer outages, delimited by zero-outage intervals. For each event, we estimate outage propagation and recovery rates at the county level, modeling them as a spatio-temporal stochastic process conditioned on weather covariates. These rates provide interpretable, region-specific resilience metrics that enable comparative analysis of system vulnerability and robustness.

To facilitate predictive capabilities, we construct a regression-based proxy for the propagation rate using meteorological variables from NOAA's High-Resolution Rapid Refresh (HRRR) dataset. This model identifies atmospheric conditions most associated with rapid outage growth, providing insights for anticipatory planning and resource allocation.

The dataset includes 20 hurricanes—ranging from Category 1 to Category 5—that primarily occurred between July and October (with Hurricane Nicole in November 2022). Since hurricanes often impact multiple states with heterogeneous effects, each state–hurricane pair is treated distinctly, resulting in 29 spatio-temporal observations. Our results support both statistical inference and engineering decision-making, offering a data-driven approach for enhancing power system resilience under escalating climate risks.

MB62

Building B Level 4 B402

Advanced High-Dimensional Data Analytics for Interconnected Systems

Invited Session

Quality, Statistics and Reliability

Chair: Elif Konyar, Georgia Institute of Technology, Atlanta, GA, 30318, United States

Co-Chair: Mohammad Bisheh, Georgia Institute of Technology, Atlanta, GA, United States

1 - Active Multi-Modal Data Fusion for Physiological Signals

Yue Zhao, Rensselaer Polytechnic Institute, Troy, NY, United States, Yinan Wang

Exploring human cognitive states through multi-modal physiological signals from multiple sources is increasingly significant in recent research. Machine learning methods often require large datasets, however, collecting such data can be costly and time-consuming. Multi-modal data fusion integrates information from diverse sources and offers a promising solution by enabling a more comprehensive understanding. However, fusing multi-modal data can usually include misleading information due to noise and data quality issues. Additionally, multi-modality datasets often contain redundant information, so it is essential to identify a subset of modalities with complementary information along with high-quality and informative data samples to improve the performance of the ML method. There are two main challenges: (1) which combination of modalities will benefit the downstream task the most, and (2) how to evaluate the informativeness and quality of data samples. We proposed a bi-level Active Multi-Modal Data Fusion framework that integrates active learning with contrastive learning to prioritize high-quality, impactful data points (lower-level optimization), while employing MaskConnect for gradient-based optimization to select the most relevant modality combinations and reduce redundancy in the dataset (upper-level optimization). Our proposed framework is validated using multi-modal physiological data collected from pilots who were tested by the simulated Instrumented Landing System (ILS), aiming to classify the level of difficulty of flight conditions. The result indicates that our framework can select the combination of modalities with complementary information and high-quality data samples to improve performance.

2 - O-Disco: Online Disentangling Signal Restoration and Corruption Identification

Deyi Kong, University of Minnesota Twin Cities, Minneapolis, MN, United States, Yifeng Wang, Shancong Mou, Jianjun Shi

The DISCO (DIsentangling Signal Restoration and Corruption Identification) method decouples the data into latent low-rank and sparse components. However, since the latent low-rank component corresponds to the nuclear norm, it requires loading the entire dataset during the optimization process. This prevents the application of DISCO to large datasets.

In this work, we introduce an Online DISCO that can process one data point or a small batch at a time, which is more efficient in terms of time and storage. We use a Frobenius norm formulation of the nuclear norm in the optimization problem, which is further solved using a block coordinate decomposition method. Simulations using the benchmark dataset BTAD verify the efficiency of O-DISCO.

3 - A Dynamic Ensemble Method for Spatio-Temporal Wind Energy Forecasting

Jiaxiang Ji, Rutgers University, Piscataway, NJ, United States, Feng Ye, Travis Miles, Ahmed Aziz Ezzat

The reliable and cost-effective operation of wind farms depends heavily on accurate wind energy forecasts across multiple look-ahead horizons, ranging from a few minutes to several days. There is an active literature on leveraging the advancements in data science to harness the multi-source information collected at, or in proximity, to the wind farms in order to make improved spatio-temporal wind energy forecasts. A central challenge in this literature is that methods optimized for short-term forecasting often perform poorly at longer horizons, and vice versa. While using separate models for different horizons is a potential solution, it can lead to discontinuities in forecasts across time and space, limiting their utility for downstream decision-making (e.g., issuing space-time scenarios with actionable uncertainty). To address this, we propose an ensemble framework that integrates two (or more) spatio-temporal models, each tailored to a specific forecast horizon. These models largely share similar structures—resulting in highly correlated outputs—but differ in their input features and parameterizations. The ensemble framework captures spatial, temporal, and cross-model correlations, and incorporates a dynamic weighting scheme that adapts to each model's changing predictive strength over time. Crucially, the method preserves probabilistic uncertainty and ensures continuous forecasts across different temporal scales. Applied to real-world wind data from the U.S. Mid-Atlantic region—an area of significant offshore wind energy potential—our approach achieves considerable improvements over standard spatio-temporal and time series forecasting methods.

4 - Improving Prediction Accuracy in Differential-Equation Based Models Using Sparse Real Observations

Soumik Sarkar, Florida State University, Tallahassee, FL, United States, Himanshu Balhara, Tapan Ganatma Nakkina, Raghav Gnanasambandam

Physics-based models, in the form of differential equations (DEs), play a vital role in modern science and engineering. These models support important decision-making, such as designing process parameters with additive manufacturing (AM), deciding policies with climate forecasting, and determining appropriate drug doses in pharmacology. However, discretization-based solvers, such as the Runge-Kutta Method, the Finite Element Method, and the Finite Volume Method, are often inaccurate in solving DE-based models. The noise in the measurement systems further contributes to the discrepancy between the solvers and real-world observations. Data assimilation techniques, such as traditional Kalman filter-based and Bayesian methods or modern neural network-based approaches, are widely prevalent to update the solution estimates of solvers with real-world observations. While the traditional methods fail to account for the inaccuracies of the solver, neural network-based approaches do not provide crucial uncertainty estimates. This work develops a novel Bayesian data assimilation method that accounts for the inaccuracies in solving the DE-based model. The case study shows accurate predictions, with uncertainty estimates, of the sparse spatio-temporal thermal data in the Directed Energy Deposition (DED) AM process by utilizing the heat transfer model. The proposed method also provides thermal gradients and cooling rates that are necessary for downstream microstructural predictions critical to quality control in metal AM.

5 - Iot Event Data Monitoring and Prediction Considering Within- and Between-Event Correlations

Zihan Yu, Tsinghua University, Beijing, China, People's Republic of, Kaibo Wang

In the era of Industry 4.0, with significant advancements in data acquisition, Internet of Things (IoT) systems are now capable of capturing a vast array of complex and diverse event data. These data often highlight sudden system anomalies that occur amidst stable signals. To address the correlations within and between events, we introduce a novel nonparametric Hawkes process model with a doubly time-dependent framework (DTNHawkes) for both baseline intensity and triggering functions,

thereby transcending the limitations inherent in traditional methodologies. Furthermore, this work develops a prediction and monitoring framework based on intensity functions estimated via a novel algorithm (DTPEM-EL), which integrates penalized expectation–maximization with the Euler–Lagrange equation. Extensive simulation experiments, conducted across a variety of scenarios, along with case studies utilizing real-world wind turbine data, have validated the efficacy of the proposed model and estimation algorithm.

6 - An Explainable AI-Driven Framework of Optimal Data Selection for Prognostics of Complex Systems

Sara Kohtz, Binghamton University, Binghamton, NY, United States, Shooq Nimri

Precise analysis and prognostics of complex systems has become an imperative research area in the field of reliability engineering. In particular, an optimal sensor network can enhance monitoring of high-risk systems across many fields. These include but are not limited to electric motors, power grids, autonomous vehicles, large-scale manufacturing, and many other concentrations. This study aims to develop a multi-fidelity modeling and data fusion framework, which can integrate data from numerous sources across different scales to achieve optimal system analysis. Within high-risk systems, there is a growing need for accurate prognostic tools that can utilize heterogeneous data from varying sources. Generalizing this framework requires techniques that optimize signals processing and sensor placement on a fundamental level. The essentials of this study lie within the scope of data selection for machine learning and data fusion models. However, the challenge of efficient data selection has yet to be formulated in the context of optimal design and sensor placement. Specifically, the proposed method can simultaneously determine the sensor network while optimizing the predictor or monitoring system. This is accomplished by using “proxy” models for data selection, and explainable AI for model transparency. The global and local explainability are provided via SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) analyses, respectively. Various search algorithms are implemented to solve the generally applicable mathematical formulation, which utilizes predictor accuracy as the fitness function. Overall, this proposed method is demonstrated on fault detection of electric motors and shows potential for implementation within impactful complex systems.

MB63

Building B Level 4 B403

Reliable Decision-Making

Contributed Session

Contributed

Chair: Yilin Wang, Simon Fraser University/Central University of Finance and Economics, Burnaby, BC, N/A, Canada

1 - Fast and Conditionally Optimal Computation of Two-Terminal Network Reliability

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

Assessing network reliability is critical for the robust design of various networked systems, but computing two-terminal (source-target) reliability is an #P-complete problem. We introduce dynamic programming (DP)-based algorithms for efficiently evaluating connectivity reliability in networks, addressing both tree-like and general topologies.

For tree-like networks, we present an asymptotically optimal DP method with linear time complexity by leveraging topological. For general networks satisfying a contraction condition ($L < 1$ where L denotes the expected reliability contribution from each node's neighbors), we develop an iterative fixed-point algorithm that guarantees convergence with bounded error, enabling polynomial-time solutions. In cases where $L \geq 1$, we develop iterative DP combined with pruning techniques to efficiently approximate reliability, particularly for large networks.

Numerical experiments on diverse networks demonstrate the effectiveness of the proposed methods, offering practical tools for scalable and robust network design.

2 - Nonparametric and Recurrent Survival Models for User Retention

Chanbeom Hur, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Korea, Republic of, Sungil Kim

Measuring user retention is essential across various industries, particularly in contexts where user engagement data are available instead of explicit churn events. Rolling retention, also referred to as bounded retention, is widely adopted but exhibits significant limitations, notably its strong dependence on the observation window length and the absence of theoretical analysis. We derive the statistical properties of rolling retention, establish confidence intervals, and propose two novel retention metrics based on survival analysis. The first metric is a nonparametric estimator of the time-to-churn distribution that assumes a single churn event per user. It incorporates rolling retention as a special case and improves the accuracy of estimating the true distribution. The second metric adopts a model-based approach that accommodates recurrent churn events. This framework identifies the time-varying features that influence churn probability and quantifies the strength of these influences. In addition, it enables dynamic, user-specific retention estimates as additional data become available over time. We demonstrate the utility of the proposed metrics through simulations and real-world applications.

3 - Improving Legacy Industrial Optimization Applications with Benchmarking

Jacob Urlick, NetJets, Columbus, OH, United States, Evelyn Arrey, Ramiro Rodriguez Buno, Anil Sharma, Ted Allen

Software applications using MIP optimization to solve business problems are common in industrial settings. Despite the prevalence and importance of these systems, there is limited available research on how to improve the performance of legacy systems, as opposed to designing performant new systems. In practice, the costs and risks associated with designing new systems can outweigh the potential benefits associated with implementing a new research technique. We propose a novel framework for improving the performance of legacy MIP systems without substantially altering the model or implementing complicated custom enhancements, and we present two case studies showing how this approach may be used on both a standard academic benchmark problem set and to improve a real-world legacy optimization application regarding the scheduling of crewmembers at a fractional ownership airline.

4 - Individuals Journey Control Charting for Monitoring Entities with Reference Groups Applied to Cybersecurity&Rsquo;S Vulnerabilities

Mohammad Abdullah, The Ohio State University, Main Campus, Columbus, OH, United States, Theodore Allen

This study introduces a novel framework for individualized risk monitoring in cybersecurity using evolving probability estimates. The proposed method, called the Individuals Journey Probability (IJP) Chart, extends Statistical Process Control (SPC) to track the changing likelihood of critical events for each entity over time. To handle non-stationary environments, the approach incorporates Conditional Statistical Control, dynamically adjusting control limits based on contemporaneous reference groups. Machine learning models are employed to periodically update the estimated probabilities for each cybersecurity vulnerability. The methodology is applied to the prediction of whether specific vulnerabilities will be included in the U.S. CISA Known Exploited Vulnerabilities Catalog. The IJP Chart effectively identifies vulnerabilities with abnormal probability trajectories relative to peer groups, providing early indicators for proactive intervention. These results demonstrate the value of probabilistic control charting for real-time, entity-focused decision support in cybersecurity.

5 - Monitoring Covariance Shifts in Streaming Multichannel Profiles

Davide Forcina, University of Naples Federico II, Naples, Italy, Christian Capezza, Antonio Lepore, Biagio Palumbo

Modern industrial systems continuously produce multichannel profile data for real-time process monitoring and fault diagnosis. Although most existing techniques focus on detecting shifts in the process mean, changes in the covariance structure are equally critical, as the process behavior depends on the interactions among multiple variables.

In this work, we use functional graphical models to capture conditional dependencies in multichannel profile data, where high dimensionality and sparsity exacerbate model estimation and interpretability.

The proposed approach employs penalized likelihood ratio tests with adaptive penalty levels to detect various types of covariance changes.

Additionally, a diagnostic procedure based on change-point detection identifies which relationships have shifted. Extensive simulations and a real-world application

involving multi-channel temperature profile monitoring confirm the proposed method's effectiveness over competing methods already appearing in the literature.

Acknowledgments

The research activity of C. Capezza, D. Forcina, and A. Lepore was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1551.11-10-2022, PE000000004).

The research activity of B. Palumbo was carried out within the MOST - Sustainable Mobility National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1033.17-06-2022, CN000000023).

This work reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

6 - Towards Interpretable Maintenance Decisions: Condition Based Maintenance Optimization with Explainable Reinforcement Learning

Stephane Barde, Hanyang University ERICA, Ansan-si, Korea, Republic of

This study investigates condition-based maintenance (CBM) in systems subject to continuous-time stochastic degradation, monitored through periodic inspections using advanced sensors—such as vibration analysis, acoustic emission, oil diagnostics, and temperature measurements. Maintenance actions are determined based on real-time assessments of system health. Optimizing CBM strategies for multi-component systems presents significant challenges due to intricate economic, stochastic, and structural interdependencies among components. Focusing on k-out-of-n systems with economic dependence—a common structure in load-sharing mechanisms, wind turbines, and aircraft systems—we model the group CBM strategy as an average reward Markov Decision Process (MDP). In economically dependent systems, group replacement policies are typically more cost-effective than individual interventions. To analyze and interpret the resulting optimal CBM policy, we propose a novel approach grounded in explainable reinforcement learning (XRL), employing reward decomposition techniques. This methodology offers unique insights into the trade-offs between different reward types by decomposing the overall average reward into component-wise contributions. This enables detailed sensitivity analyses, identifying the reward components with the highest influence, and supports clearer comparisons with benchmark methods. The XRL framework enhances interpretability by revealing the rationale behind the optimal decisions relative to heuristic alternatives. Numerical experiments validate the method's ability to facilitate concise sensitivity assessments and benchmark comparisons, underscoring the practical value of the proposed approach.

7 - Congestion-Based Repair Policy for a Failure-Prone Service System with Strategic Customers

Yilin Wang, Central University of Finance and Economics, Beijing, China, People's Republic of, Jinting Wang, Lingjiao Zhang, Zhe Zhang

This study examines the decision-making interaction between a service provider adopting a congestion-based repair policy and strategic customers in a failure-prone M/M/1 queueing system. The server's lifetime is exponentially distributed, and a repair starts immediately upon server's breakdown. The repair rate is adjustable: the service provider employs a high repair rate (with a high cost) if the number of waiting customers reaches a threshold; otherwise, a low repair rate (with a low cost) is adopted. We aim to find the optimal congestion-based repair policy by taking customers behavior into account under two different information scenarios. To begin with, the Nash equilibrium strategies and socially-optimal strategies in the unobservable case are derived. Both follow-the-crowd and avoid-the-crowd behaviors are found coexisted in the customer's equilibrium joining strategy. Two special models, classic repair model (when the threshold approaches 0) and delayed repair model (when the low repair rate approaches 0), are discussed extensively. The classic repair policy maximizes throughput but incurs the highest costs, while delayed repair minimizes costs at the expense of throughput. The proposed congestion-based repair strategy balances these trade-offs, achieving intermediate throughput and cost levels. Notably, it can increase profits by up to 34.4% compared to classic repair, with its effectiveness amplified under high-cost scenarios. By comparing the unobservable case with the almost unobservable counterpart, we demonstrate that hiding server state information when prices are low and disclosing server information when prices are high can increase profit for service provider, but at the expense of reducing social welfare.

MB64

Building B Level 4 B404

Recent Advances in Behavioral Operations

Invited Session

Behavioral Operations Management

Chair: Ignacio Rios, The University of Texas at Dallas, 632 Northhill Dr, Richardson, TX, 75080, United States

1 - Designing Effective Fundraising Campaigns: Behavioral Insights

Pramit Ghosh, The University of Texas at Dallas, Dallas, TX, United States, Ignacio Rios, Anyan Qi

This study examines how fundraisers maximize donations under budget constraints, focusing on financial incentives (matching or gift unlock) and solicitation methods (simultaneous or sequential). Using a game-theoretic model and laboratory experiments, we provide insights into effective fundraising strategies and behavioral deviations in donor responses.

2 - AI-Powered Philanthropy: Effects on Volunteer Productivity

Vinit Tipnis, Carlson School of Management, University of Minnesota, Minneapolis, MN, United States, Eunae Yoo, Gloria Urrea, Fei Gao

Nonprofit organizations are incorporating AI to enhance volunteer performance. Given that volunteers are motivated by generating impact and altruism, the effect of AI in the volunteering context is an open question. We answer this using data from an online volunteering platform that implemented an AI tool that complements volunteers' work.

3 - Improving Food Security with Scheduled Neighborhood Grocery Delivery

Yeonjoo Lee, University of Minnesota, MINNEAPOLIS, MN, United States, Karen Donohue

Food security is a pressing concern in the US, with 13.5% of its population identified as food insecure. While online grocery delivery has recently surfaced as a potential solution, the high delivery fees and markup pricing typical of most on-demand delivery models (e.g., Instacart) pose critical cost barriers for low-income populations. This study examines how a novel grocery delivery model, called "scheduled neighborhood delivery," which consolidates nearby grocery orders and delivers on a weekly schedule, can be a better alternative for the food security problem. In collaboration with a startup grocery delivery platform utilizing this service, we investigate the potential operational and customer well-being benefits of this new delivery service. First, using work measurement data, we estimate the magnitude of operational efficiency increase allows scheduled neighborhood delivery to charge lower the shipping fees. Next, we conduct controlled experiments explore the impact of alternative online grocery delivery designs on consumer purchase behavior, with a focus on healthy food selection. Results show the potential for increased healthy food choices under the scheduled neighborhood delivery model compared to the on-demand delivery model, especially for planned grocery purchases.

4 - Reference-Dependent Effects in Volunteering: Expected vs. Experienced Effort

Gabriel Pensamiento, University of Pittsburgh Katz Graduate School of Business, Pittsburgh, PA, United States, Mahyar Eftekhari, Leon Valdes

Humanitarian and nonprofit organizations often rely on volunteer labor to accomplish different tasks. We design an incentivized laboratory experiment to study how someone's volunteering experience being above or below effort expectations affects their decision to volunteer again. In doing so, we study the role of reference-dependent preferences in a not-for-profit setting, which is not well understood in the literature. We also aim to shed light on which responses may be unique to (or more pronounced in) a volunteering setting by comparing our results against a similar task in a for-profit context.

MB65

Building B Level 4 B405

Emergent Intelligence: AI in Action

Invited Session

Artificial Intelligence

Chair: Juan Qin, University of Science and Technology of China, SEATTLE, WA, N/A, United States

Co-Chair: Kai Sun, University of Florida, Sunderland, MA, United States

1 - The Impact of Loyalty Program on Gig Workers: Evidence from a Natural Experiment**Jiaxin Lei, Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Gen Li, Jing Wang, Yao Zhao**

The rapid growth of on-demand delivery platforms has transformed labor dynamics within the gig economy, introducing new challenges in motivating a flexible, independent workforce. While prior literature has focused on incentive schemes such as surge pricing, penalties, and rating systems, we investigate the causal effect of a loyalty program that enables gig workers to earn virtual points by completing orders and redeem them for rewards. Leveraging a unique dataset from a major Asian online delivery platform, we assess how the program influences riders. We exploit the staggered rollout of the loyalty program across riders by comparing behavioral changes between riders who receive loyalty points and those who do not. The findings suggest the loyalty program leads to longer work durations and more delivery order volume. Further analysis reveals simply allocating virtual points does not produce a significant effect on riders. Instead, the positive overall impact is primarily driven by the redemption process. More interestingly, riders shift their behavior by delivering more small, lightweight, and short-distance orders. The pursuit of greater order quantity has led riders to accept more small orders, thereby decreasing their earnings per unit of time. While riders are working for longer durations, their tendency to take on smaller orders has not yielded higher total earnings. Compared to full-time and experienced riders, the program has a greater impact on elite freelance and inexperienced riders. These findings deepen our understanding of loyalty programs' effects on gig workers and offer insights for platforms aiming to boost productivity through flexible incentives.

2 - Reimagining Digital Markets: Platform Dominance vs. Open-Network Models**Shailender Joseph, University of Texas at Dallas, Richardson, TX, United States, Kai Sun, GANESH JANAKIRAMAN, Vijay Mookerjee**

In this work, we contrast the traditional “walled-garden” architecture epitomized by Amazon – where one firm owns the marketplace, controls the buyer interface, and levies non-negotiable commissions – with the Open Network for Digital Commerce (ONDC), an open protocol that functions like a shared system of digital railroads. Networks like ONDC decouple the buyer-side interactions and seller-side functions, making every seller discoverable to every buyer and preventing any single intermediary from sitting at the center of each transaction. Because this network-based paradigm is still in its infancy, its economic impact remains largely uncharted. Our work constructs a game-theoretic framework to quantify how migrating from a closed platform to an open network reshapes seller surplus, buyer welfare, and intermediary revenues and pinpoints the conditions under which each market design yields superior efficiency for e-commerce participants.

3 - Genai in Creative Production: Does It Narrow the Gap Between High- and Low-Skill Creators?**Zhen Yan, University of Texas at Dallas, Richardson, TX, United States, Jianqing Chen, Srinivasan Raghunathan**

The rapid development and widespread adoption of AI are transforming many industries, including creative arts and software—fields traditionally regarded as strongholds of human

creativity. A central question that remains debated is whether AI will democratize creative production by narrowing the output gap between high-skilled and low-skilled creators. This study investigates how AI adds value to the creative process and how it influences the product-value gap across skill levels. We model creative products as having two key attributes: ideation and execution. AI affects both through two distinct mechanisms: automation, which substitutes for certain execution tasks, and augmentation, which enhances creators' ideation capabilities. We find that AI consistently improves product value for creators at all skill levels. However, the magnitude of this improvement varies across creators: AI tends to widen the product-value gap when the high-skilled creator is ideation savvy (i.e., holds a greater advantage in ideation than execution) or when the high-skilled creator is execution savvy, but the automation is relatively weak. In other scenarios, AI may instead narrow the gap. These outcomes are driven by shifts in effort allocation, as AI both reduces the need for human effort in execution and increases the effectiveness of ideation. As a result, creators reallocate effort from execution to ideation. The widening of the gap may result from growing disparities in both attributes—or from ideation alone—or from increased complementarity between ideation and execution.

4 - Beyond Taste: Leveraging LLMs to Mitigate Bias in Online Reviews**Di Yuan, University of Illinois Urbana-Champaign, Urbana, IL, United States**

Online reviews heavily influence what we buy, especially for experience goods. But these reviews often contain hidden biases that can unfairly skew our perception of a product. Our study zeroes in on taste-based bias, where a reviewer's personal preferences—rather than objective quality—cloud their judgment. We're using movie reviews to explore how individual tastes can overshadow objective assessments of cinematic quality.

We've developed a new approach that uses **large language models (LLMs)** to pinpoint and rectify this taste-based bias. Our method involves fine-tuned transformer models that learn to differentiate between genuine critiques of quality and evaluations driven purely by personal preference. This is achieved by analyzing linguistic patterns, the attribution of sentiment, and the specific focus of the content. We design this method to pinpoint reviews where personal taste disproportionately impacts ratings and subsequently generate bias-corrected versions that emphasize objective product attributes.

Preliminary results indicate a significant enhancement in review quality metrics and a reduced variance in product ratings following bias correction. This study contributes to the burgeoning field of automatic content moderation and offers practical tools for platforms seeking to improve the reliability of their review systems. Furthermore, our findings have important implications for the design of recommendation systems, market research, and the analysis of consumer behavior in digital marketplaces.

MB66

Building B Level 4 B406

Exploring and Maintaining Teaching Stream Careers in Academia

Panel Session

INFORMS Committee on Teaching and Learning

Co-Chair: Craig Calvert, University of Connecticut, Storrs, CT, United States

Co-Chair: Yaren Bilge Kaya, Columbia University, New York, NY, United States

1 - Panelist**Gabriela Gongora-Svartzman, Northeastern University, Miami, FL, United States****2 - Panelist****Yaren Bilge Kaya, Columbia University, New York, NY, United States****3 - Panelist****Mariah Magagnotti, Clemson University, Erie, PA, United States****MB67**

Building B Level 4 B407

Generative Models for Statistics and OR

Invited Session

Data Mining

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

1 - Guided Diffusion for Statistical Inference**Yuchen Wu, University of Pennsylvania, Philadelphia, PA, United States**

Diffusion models benefit from instillation of task-specific information into the score function to steer the sample generation towards desired properties. Such information is coined as guidance. For example, in text-to-image synthesis, text input is encoded as guidance to generate semantically aligned images. Proper guidance inputs are closely tied with the performance of diffusion models. A common observation is that strong guidance promotes a tight alignment to the task-specific information, while reduces the diversity of the generated samples. In this paper, we provide the first theoretical study towards the influence of guidance on diffusion models in the context of Gaussian mixture models. Under mild conditions, we prove that incorporating diffusion guidance not only boosts prediction confidence but also diminishes distribution diversity, leading to a reduction in the differential entropy of the output distribution. Our analysis covers the widely used DDPM and DDIM sampling schemes, and leverages comparison inequalities in differential equations as well as the Fokker-Planck equation that characterizes the evolution of probability density function, which may be of independent theoretical interest.

2 - Diffusion Transformers for Imputation: Statistical Efficiency and Uncertainty Quantification**Zeqi Ye, Northwestern University, Evanston, IL, United States, Minshuo Chen**

Imputation methods play a critical role in enhancing the quality of practical time-series data, which often suffer from pervasive missing values. Recently, diffusion-based generative imputation methods have demonstrated remarkable success compared to autoregressive and conventional statistical approaches. Despite their empirical success, the theoretical understanding of how well diffusion-based models capture complex spatial and temporal dependencies between the missing values and observed ones remains limited. Our work addresses this gap by investigating the statistical efficiency of conditional diffusion transformers for imputation and quantifying the uncertainty in missing values. Specifically, we derive statistical sample complexity bounds based on a novel approximation theory for conditional score functions using transformers, and, through this, construct tight confidence regions for missing values. Our findings also reveal that the efficiency and accuracy of imputation are significantly influenced by the missing patterns. Furthermore, we validate these theoretical insights through simulation and propose a mixed-masking training strategy to enhance the imputation performance.

3 - Flow Generative Models for High-Dimensional and Multi-Modal Statistical Sampling**Dongze Wu, Georgia Institute of Technology, Atlanta, GA, United States**

Sampling from high-dimensional, multi-modal statistical distributions remains a fundamental challenge across domains such as statistical Bayesian inference and physics-based machine learning. In this talk, we present Annealing Flow (AF), a method built on Continuous Normalizing Flow (CNF) for sampling from high-dimensional and multi-modal distributions. AF is trained with a dynamic Optimal Transport (OT) objective incorporating Wasserstein regularization, and guided by annealing procedures, facilitating effective exploration of modes in high-dimensional spaces. Compared to recent NF methods, AF greatly improves training efficiency and stability, with minimal reliance on MC assistance. We demonstrate the superior performance of AF compared to state-of-the-art methods through experiments on various challenging distributions and real-world datasets, particularly in high-dimensional and multi-modal settings. We also highlight AF potential for sampling the least favorable distributions.

4 - Interactive Decision-Making Using Autoregressive Generation**Hongseok Namkoong, Columbia University, New York, NY, United States**

AI agents interacting with the real-world have to grapple with a perpetual lack of data in ever-changing environments. Interactive decision-making requires going beyond knowledge distillation: an intelligent agent must comprehend its own uncertainty and actively gather information to resolve it. Despite exhibiting impressive capabilities in basic knowledge work, state-of-the-art AI systems struggle to articulate their own uncertainty, e.g., OpenAI recently noted its latest agentic system “DeepResearch often fails to convey uncertainty accurately”.

This talk covers a series of recent works that tackle the central challenge of uncertainty quantification in natural language-based interactive decision-making problems. Instead of modeling latent environment parameters, we view uncertainty as arising out of missing future outcomes and quantify it through autoregressive sequence generation—iteratively predicting the next outcome given the past. By adapting to new information through in-context learning rather than cumbersome posterior inference, our approach seamlessly scales to problems involving unstructured data, e.g., adaptive student assessment involving text and images. Formally, we establish a reduction from online decision-making to offline next-outcome prediction, enabling us to leverage the enormous datasets and computational resources dedicated to improving sequence prediction for interactive decision-making tasks.

5 - Finite-Time Convergence Guarantees for Masked Diffusion Models**Yingbin Liang, The Ohio State University, Columbus, OH, United States, Yuchen Liang, Renxiang Huang, Lifeng Lai, Ness Shroff**

Masked diffusion models (MDMs) have recently emerged as a powerful class of generative models for discrete tasks such as language generation, code synthesis, and molecular design, by iteratively denoising masked tokens. Despite their impressive empirical performance, a rigorous theoretical understanding of their convergence behavior and error dynamics remains largely unexplored. In this work, we provide the first finite-time error bounds and convergence guarantees for MDMs. We first analyze the convergence of the forward process, where we address a key challenge posed by the degeneracy of the masking stationary distribution in the KL divergence by introducing a novel surrogate initialization distribution. We then establish the first convergence guarantees for both the ϵ -leaping and uniformization samplers for MDMs, showing improved rates compared to those using uniform rate matrices. Furthermore, under mild assumptions, we establish convergence without the need for early stopping. Our analysis introduces several new technical tools to address challenges unique to MDMs, including a Jensen-type argument for bounding forward process convergence, novel techniques for bounding masking score functions, and a non-divergent upper bound on the score near initialization that removes the need of early-stopping.

MB68

Building B Level 4 B408

Evaluation and Scalability of LLM Prompting

Invited Session

Data Mining

Chair: Hongyang Zhang, Northeastern University, Boston, MA, United States

Co-Chair: Hongseok Namkoong, Columbia University, New York, NY, United States

1 - Past Experience Accelerates LLM Reasoning**Bo Pan, Emory University, Atlanta, GA, United States, Liang Zhao**

Allocating more compute to large language models (LLMs) reasoning has generally been demonstrated to improve their effectiveness, but also results in increased inference time. In contrast, humans can perform tasks faster and better with increased experience and practice. Hence, we aim to investigate the question: Can LLMs also become faster at reasoning through recurrent practice on relevant tasks, and if so, how?

We first formalize the problem setting of LLM reasoning speedup systematically in the dimensions of task relevancy and compute budget calculation. We then propose framework to implement and benchmark such reasoning speedup behaviour based on adaptive compute allocation and memory mechanisms. We further conduct comprehensive experiments to benchmark such behaviour across different question similarity levels, memory methods, and reasoning methods. Results show that LLMs can generally reason faster with practice, achieving up to a 56% reduction in compute cost when equipped with appropriate memory and reasoning methods.

2 - Label-Invariant Sharpness Regularization to Mitigate Grokking in Mathematical Reasoning

Zhenshuo Zhang, Northeastern University, Boston, MA, United States

Mathematical reasoning is a central aspect in the evaluation of large language models. A phenomenon observed in modular arithmetic tasks, a basic math problem, is grokking, where training performance converges to perfect accuracy but test performance does not, until after an extended period. Understanding grokking and, more generally, the generalization of LLMs for math reasoning remains a mystery. This paper examines the presence of invariant structures in math reasoning, including arithmetic and graph-algorithmic tasks. We find that the sharpness of the loss surface, as defined by its Hessian trace, is much lower along label-invariant input directions than a random input direction. Let g denote a label-invariant transformation and x denote an input. A label-invariant direction refers to a direction that moves along the augmentation of x via g , minus x . Thus, the "label-invariant Hessian trace" provides a more precise sharpness measure for the loss landscape of math reasoning tasks. Motivated by this finding, we design a new regularization method to penalize the Hessian trace of the loss surface along label-invariant input directions. First, we interpolate between x and an augmented version of x using g . Second, we compute the gradient at the interpolating point while adding noise to the model weights. We also provide a generalization bound that depends on a Hessian distance measure for learning invariant function classes to tie this together. For six text-based arithmetic and graph-algorithmic tasks, our approach improves the test accuracy of LLMs by 16.5% and by 69% on out-of-domain samples.

3 - AI-Powered Recruitment Evaluation: from Intuition-Based to Data-Driven Hiring Decisions with A Generative Agent

Jaeyeon Yoon, Yonsei, Seoul, Korea, Republic of, Young Il Oh, Hyunsoo Yoon

Traditional hiring relied heavily on human intuition, resulting in manual and subjective decisions with limited explainability. We developed a generative AI assistant that analyzes resumes, summarizes competencies, verifies facts, assesses job and culture fit, recommends roles, and generates tailored interview questions. Comparing recruiter choices before and after AI use highlights the need for structured, explainable support in hiring.

4 - Scaling up Inference via Gradient Estimation

Hongyang Zhang, Northeastern University, Boston, MA, United States

This paper introduces an algorithm to select demonstration examples for in-context learning of a query set. Given a set of n examples, how can we quickly select k out of n to best serve as the conditioning for a downstream task? This problem has broad applications in prompt tuning and chain-of-thought reasoning, to name a few. Since model weights remain fixed during in-context learning, previous work has sought to design methods based on similarity scores measured in the input embeddings. This work proposes a new approach based on gradients of the model output taken in the input embedding space. Our approach estimates model outputs through a first-order approximation using the gradients. Then, we apply this estimation to multiple randomly sampled subsets. Finally, we aggregate the sampled subset outcomes to form an influence score for each demonstration, and select k most relevant examples. This procedure only requires pre-computing model outputs and gradients once, leading to a linear-time algorithm relative to model and training set sizes. Extensive experiments across various LLMs and datasets validate the efficiency of our approach. We show that the gradient estimation procedure yields approximations of full inference with less than 1% error across six datasets. This allows us to scale up subset selection methods that would otherwise run full inference by up to 37.7 times on LLMs with up to 34 billion parameters, and outperform existing selection methods based on input embeddings by 11% on average.

MB69

Building B Level 4 B409

Design, Behavior, and Performance Topics in the Knowledge Economy

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Sina Khorasani, Vanderbilt University, Nashville, TN, United States

Co-Chair: Konstantinos Stouras, UCD Michael Smurfit GSB, Dublin, Ireland

1 - A Machine Learning Framework for Project Performance Prediction, Interpretation, and Inference

Xiaochen Kev Gao, University of California, San Diego, La Jolla, CA, United States, Sina Khorasani, Vish Krishnan, Lakshminarayana Nittala, Siqi Wang

In this paper, we offer a data-driven AI-based analytic pipeline to predict, interpret, and infer the nature and severity of cost or schedule overrun a project will experience. We integrate large language models (LLMs) in the pipeline for data augmentation and prediction, and operate on unstructured data in addition to tabular data. Our analysis and results underscore the need for the government and other enterprises to collect and utilize the data for predictive and causal analysis and to develop more nuanced approaches and incentives to tackle cost and schedule overruns in real-world project settings.

2 - Enhancing the Entrepreneurial Process with Large Language Models

Karthik Ramachandran, Georgia Institute of Technology, Atlanta, GA, United States, Nitin Joglekar

We examine how Large Language Models (LLMs) support the entrepreneurial process by facilitating the creation and exploration of opportunities. An experimental study reveals both their potential and challenges, and shows how different training approaches influence the effective use of LLMs in entrepreneurial contexts.

3 - How Do Loyalty Programs Affect Multichannel Bookings? An Empirical Investigation

Shuaishuai Yang, UCL School of Management, London, United Kingdom, Antoine Feylessoufi, Ersin Korpoglu

With the increased market power of online travel agencies (OTA), the hotel industry faces a new and important trade-off when working with OTAs: Potentially increasing sales at the expense of sizeable commission fees. In response, major hotel groups turn to loyalty programs, giving discounts to loyalty members for booking through their direct channels. In this paper, we investigate how successful loyalty discounts are in their goal to sway customers from OTAs to direct channels. Using a unique dataset obtained from a major hotel group, we empirically investigate how loyalty discounts affect customers' booking channel choices. We utilize a quasi-natural experiment based on an exogenous change of the loyalty discount in a subset of hotels in our dataset to construct a difference-in-difference model. We find that a larger loyalty discount raises bookings from the direct channel with loyalty membership, lowers bookings from the direct channel without loyalty membership, and has no significant effect on OTA bookings. This suggests that a substitution effect only occurs within the direct channel and, surprisingly, not cross-channels. Our results indicate that discount-based loyalty memberships do not have the intended effect of swaying customers from OTAs toward direct channels. Instead, directchannel customers utilize the loyalty program more, which may end up lowering hotel revenues. Therefore, our study suggests that the hotel industry should carefully assess the effectiveness and the design of their loyalty programs by factoring in their multichannel setting.

4 - On Repeat: Iteration as a Key to Innovative Performance?

Christian Jost, University of Augsburg, Augsburg, Germany, Evgeny Kagan, Sebastian Schiffels, Tobias Lieberum

Motivated by the rise of iterative project management, we examine how workflow -- iterative or sequential -- affects innovative behavior and performance through a series of laboratory experiments. Our first experiment shows that, in an open-ended creative challenge, iterative task completion outperforms sequential task completion. In the second experiment, we show that the advantage of the iterative workflow extends to innovation settings that do not involve idea generation. A key driver of this result is that iterative work results in frequent task switching, encouraging a broader search for the best solution. Our third experiment dives deeper into the search process and shows that sequential work indeed leads to more myopic idea refinement behaviors, often resulting in suboptimal outcomes. Overall, our results suggest that the iterative workflow enhances performance across diverse innovation settings, though there are some key limitations: (1) the iterative workflow helps to achieve quick gains, but its performance advantage narrows over time. Consequently, workflow effects are stronger when balanced performance across project components is required, but weaker

when excellence in one component can offset poor performance in others. (2) The iterative workflow offers little benefit in tasks that do not require the broad exploration of solutions. (3) The iterative workflow can be harmful in settings with strong technological path dependencies.

5 - Influence Structures and Collective Action In Online Communities: Evidence from Reddit's R/place

Tian Chan, Emory University, Atlanta, GA, United States, Shizhen Chen, Anandhi Bharadwaj

Online communities frequently grapple with two fundamental challenges in collective action projects: coordinating strategy and mobilizing broad-based participation. This study posits that the structure of influence within an online community significantly impacts its efficacy in addressing these challenges. Specifically, we argue that communities with more hierarchical influence structures—where a concentrated few significantly shape others' behavior—excel at coordinating strategy but struggle with mobilizing participation. Conversely, flatter structures may face initial coordination hurdles but foster greater participation. We test these hypotheses using empirical data from Reddit's 2022 r/place event, where communities mobilize their members to construct artwork pixel-by-pixel on a shared digital canvas. Our findings reveal several key insights: Communities with more hierarchical influence structures demonstrate faster strategic decision-making, evidenced by earlier project initiation and more rapid initial progress toward their objectives. However, this early advantage shrinks as inter-community competition intensifies, increasing the imperative to mobilize marginal members. In these more demanding phases, flatter influence structures surpass their hierarchical counterparts in project progression rates. The performance improvement of flat communities is particularly pronounced when they possess a large contingent of marginal members who subscribe to community updates but do not actively participate in discussions. These findings advance our understanding of online community dynamics and offer crucial implications for the design and management of various online collective action projects, such as crowdfunding campaigns, online petitions, digital protests, and coordinated stock trades.

Monday, October 27, 1:15 PM - 2:30 PM

MC01

Building A Level 3 A301

FICO/Gurobi Optimization

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Reducing Time-to-Value with Fico Xpress Optimization Suite and Gen AI

Carlos Zetina, FICO Xpress Optimization, Québec, QC, Canada

Time to Value measures how long it takes for technology projects to begin providing business value. Three critical milestones to achieve this are model creation, validation, and deployment. In this showcase, we'll show how FICO Xpress Optimization Suite shortens Time to Value and supercharges Return on Investment (ROI) by providing a unified platform for development, validation, and deployment. We'll show how FICO's suite of products and Gen AI come together to turbocharge building and adoption of enterprise optimization applications deployable on any tech stack, be it server-less or via virtual machines, hosted on any cloud provider or on-premises infrastructure.

Hear from Rusty Burlingame from Southwest Airlines, on how Southwest is using Xpress to improve customer experience, enhance employee engagement, and streamline operations to maximize revenue and support growth.

2 - Cracking the Vault with Gurobot: Generative AI and the Future of Mathematical Modeling

Cara Touretzky, Gurobi Optimization, Santa Fe, NM, United States

Discover how Gurobi is exploring the use of generative AI to support and streamline the modeling of mathematical optimization problems. Through the lens of a hands-on museum heist challenge, we'll walk through examples of AI-assisted modeling across varying levels of complexity. We'll highlight current capabilities, acknowledge present limitations, and offer a glimpse into where we're headed next as we expand the role of AI in the modeling workflow.

MC02

Building A Level 3 A302

Advances in Quantum Computing Optimization

Invited Session

Committee's Choice

Chair: Tamás Terlaky, Lehigh University, 200 W. Packer Avenue, Bethlehem, PA, PA, 18015, United States

1 - A Quantum Central Path Method for Linear Optimization

Giacomo Nannicini, University of Southern California, Los Angeles, CA, United States

We propose a novel quantum algorithm for solving linear optimization problems by quantum-mechanical simulation of the central path. While interior point methods follow the central path with an iterative algorithm that works with successive linearizations of the perturbed KKT conditions, we perform a single simulation working directly with the nonlinear complementarity equations. This approach yields an algorithm for solving linear optimization problems involving m constraints and n variables to ϵ -optimality using, approximately, $O(\sqrt{m+n} \|R_1\|/\epsilon \text{nnz}(A))$ elementary gates, where $\text{nnz}(A)$ is the number of nonzero elements of the constraint matrix, R_1 is the norm of the optimal solution. Our scheme does not require QRAM.

2 - On Accelerating Optimization Methods by Quantum Linear Algebra

Mohammadhossein Mohammadisiahroudi, Lehigh University, Hatfield, PA, United States, Tamás Terlaky, Zeguan Wu, Pouya Sampourmahani

Solving systems of linear equations is a core component of many optimization algorithms. Quantum linear algebra techniques, especially quantum linear system algorithms (QLSAs), offer the potential to significantly accelerate these computations. One area where this promise is particularly compelling is in Quantum Interior Point Methods (QIPMs), which aim to solve large-scale conic optimization problems more efficiently by leveraging QLSAs. While QIPMs offer favorable scaling with respect to dimension, they face several challenges. These include sensitivity to problem conditioning, errors introduced by quantum tomography, and the limitations of current quantum hardware. Recent research has addressed some of these issues through various strategies: iterative refinement improves accuracy of solution, preconditioning enhances the conditioning of systems, and linear system reformulations maintains feasibility even with error at each iteration. A particularly promising development is the almost-exact QIPM framework, which is more quantum-enhanced as classical matrix-vector (mat-vec) operations are replaced with quantum mat-vec operations. This approach not only reduces noise sensitivity but also achieves *optimal dimension scaling*, leading to a clear quantum advantage for linear optimization. Additionally, in the domain of PDE-constrained optimization, which is crucial in engineering and scientific computing, QLSAs can solve the adjoint and forward equations efficiently without requiring tomography. This results in an exponential quantum speed-ups with respect to number of state variables. In this talk, I will provide an overview of these developments, discuss their implications, and outline open questions and future directions in applying quantum linear algebra to accelerate optimization.

3 - Exponential Quantum Speedup in Optimization: Theory and Practice

Xiaodi Wu, University of Maryland, College Park, College Park, MD, United States, Jiaqi Leng, Yufan Zheng

Harnessing the power of quantum algorithms to tackle optimization problems is a critical frontier with transformative potential for both theory and real-world applications. Traditional approaches to quantum speedups in optimization typically focus on accelerating specific subroutines within classical algorithms, while preserving the overall algorithmic structure and solution quality. In contrast, we introduce Quantum Hamiltonian Descent (QHD)—a fundamentally quantum optimization framework derived from the path integral of dynamical systems associated with classical optimization methods.

We establish both strong theoretical foundations and empirical evidence for the effectiveness of QHD, based on current analytical tools and available quantum hardware. Specifically, we prove that QHD achieves an exponential quantum-classical separation for a class of oracular continuous optimization problems. Beyond the oracular setting, we demonstrate that QHD can efficiently solve a family of nonconvex degree-4 polynomial optimization problems with exponentially many local minima—instances for which state-of-the-art classical solvers, including Gurobi, appear to require super-polynomial time.

To enable practical implementation, we propose a novel Hamiltonian embedding technique, and experimentally realize QHD on an analog quantum Ising simulator with over 5000 spins, solving nonconvex quadratic programming (QP) problems of up to 75 dimensions. To support broader adoption, we also release QHDOPT, a user-friendly software package built on the QHD framework, designed to empower domain experts in leveraging quantum optimization.

4 - Operator-Level Quantum Acceleration of Non-Logconcave Sampling

Lin Lin, UC Berkeley, Berkeley, CA, United States

Sampling from probability distributions with a continuous potential, is a fundamental task across physics, chemistry, biology, computer science, and statistics. However, when the potential is non-convex, the resulting distribution becomes non-logconcave, and classical methods such as Langevin dynamics often exhibit poor performance. We introduce the first quantum algorithm that provably accelerates a broad class of continuous-time sampling dynamics. For Langevin dynamics, our method encodes the target Gibbs measure into the amplitudes of a quantum state, identified as the kernel of a block matrix derived from a factorization of the Witten Laplacian operator. This connection enables Gibbs sampling via singular value thresholding and yields the first provable quantum advantage with respect to the Poincaré constant in the non-logconcave setting. Building on this framework, we further develop the first quantum algorithm that accelerates replica exchange Langevin diffusion, a widely used method for sampling from complex, rugged energy landscapes.

5 - Recent Progress in Quantum Approximate Optimization

Ruslan Shaydulin, JPMorgan Chase, New York, NY, United States

I will present recent research progress in quantum approximate optimization at the Global Technology Applied Research center of JPMorganChase

MC03

Building A Level 3 A303

Joint Session for the MCDM and GDN Sections

Invited Session

Group Decision and Negotiation

Chair: Eduarda Frej, Universidade Federal de Pernambuco, Recife, Brazil

Co-Chair: Danielle Costa Morais, Universidade Federal de Pernambuco, Recife, N/A, Brazil

1 - Multicriteria Group Decision-Making Using Flexible and Interactive Tradeoff with Partial Preference Information

Adiel De Almeida, Universidade Federal de Pernambuco, Recife, Brazil, Eduarda Frej, Jonatas Almeida, Ana Paula Cabral Seixas CostaCosta

The presentation shows a new Multicriteria Group Decision-Making (MCGDM) process based on Flexible and Interactive Tradeoff (FITradeoff), in order to deal with partial preference information. An emphasis on the conceptual decision process is given. A short view of a web-based Group Decision Support System (FITradeoff-GDSS) is given, representing a practical decision tool for aiding the solution of problems of group decision with multiple criteria. The system uses only partial and minimum preference information from different DMs, requiring less cognitive effort in a more confident way. The preference modeling process combines both decomposition and holistic evaluation. The decision process can be implemented with separate or simultaneous elicitation processes for the group of DMs. In both processes the analyst and DMs can concentrate the analysis on the potential optimal alternatives of DM and also the information of alternative rankings for each DM. Several visualizations with partial results are provided by the FITradeoff-GDSS, at each step of the preference modeling process. This allows DMs can reach a final agreement without going further in the elicitation process. Also, analytical aggregation tools can be provided considering both cardinal and ordinal aggregation approaches.

2 - Multicriteria Group Decision Process with the FITradeoff Group Decision Support System (FITradeoff-Gdss)

Wagner Farias Gouveia, Federal University of Pernambuco, Recife, Brazil, Eduarda Frej, Adiel de Almeida

This work aims to present the FITradeoff Group Decision Support System (FITradeoff-GDSS), which operationalizes the group decision process with the FITradeoff method. FITradeoff-GDSS is an interactive web-based system with several flexibility features for users to take advantage of. The preferences elicitation process of group members may be conducted either simultaneously or independently. Based on individual results analysis, there are multiple ways to address the group decision problem and reach a final solution. This can be done considering a participatory process, in which results can be analyzed and discussed considering different visualization tools provided by the GDSS. Alternatively, if no agreement is reached based on this participatory process, different aggregation tools available in the system may be applied to find a compromise solution for the group. The preferences modeling process of group members in FITradeoff-GDSS should be guided by an analyst, who must aid the DMs in the whole process. This work will present an overall description of the system, raising all flexibility features and tools applied to facilitate the group decision process.

3 - A Negotiation Model for Conflict Resolution in A Brazilian Watershed

Danielle Costa Morais, Universidade Federal de Pernambuco, Recife, Brazil, Caio Vinicius Araujo Ferreira Gomes

Water resources management involves complex processes with multiple stakeholders whose interests often conflict, leading to disputes that may manifest in various forms, such as access to water, availability, resource quality, and prioritization of use. In this context, this study proposes a negotiation model for mediating conflicts within Watershed Committees. To demonstrate the applicability of the proposed negotiation model, a simulation was conducted in the Salgado River sub-basin, located in the semi-arid region of northeastern Brazil, considering a bilateral conflict between water users and the managing authority. The defined issues included environmental education, waste management, sustainable agricultural practices, fisheries control, safe water access, and the monitoring of water use. The negotiation was structured into three stages: (i) preference elicitation, conducted using the FITradeoff method, which generates a ranking of option sets; (ii) offer exchange, carried out through successive rounds based on the obtained rankings, allowing negotiators to formulate proposals and evaluate counterproposals; and (iii) agreement definition, consolidating a compromise solution between the parties. The rankings were compared, confirming that the agreement constituted a Pareto-optimal compromise, as no alternative would yield better results for one party without compromising the interests of the other. This study highlights the proposed model as a decision support tool for negotiation in participatory management contexts, with potential applicability to a broad range of conflicts in river basins. The use of the proposed model promotes greater transparency, encourages constructive dialogue, ensures a balance of stakeholder interests, and reinforces the legitimacy of collaborative decision process.

4 - A Multicriteria Group Decision Model for Managing Risk in Urban Floods

Lucas da Silva, Universidade Federal de Pernambuco, Recife, Brazil, Danielle Morais, Marcelo Alencar, Rodrigo Ferreira, Adiel de Almeida

Fostering resilient societies against climate-driven extreme events is challenging due to uncertain and conflicting urban strategies. From a multidimensional perspective, this study undertakes a group-based decision model grounded in a multicriteria framework that integrates human, environmental, and economic criteria to evaluate urban floods under risk and uncertainty. The model accounts for bringing together decision-makers who combine their preferences and knowledge to enhance problem modeling in risky situations in a changing climate. A numerical case study supported by graphical and simulation techniques offers a comprehensive analysis, showcasing how the approach can strengthen collaborative urban adaptation strategies in effective and applicable ways.

MC04

Building A Level 3 A304

Finface - Student Paper Competition

Award Session

Finance

Chair: Renyuan Xu, Stanford University, Palo Alto, CA, United States

1 - Distributional Adversarial Attacks and Training in Deep Hedging

Guangyi He, Imperial College London, London, United Kingdom

We study the robustness of classical deep hedging strategies under distributional shifts by leveraging the concept of adversarial attacks. We first demonstrate that standard deep hedging models are highly vulnerable to small perturbations in the input distribution, resulting in significant performance degradation. Motivated by this, we propose an adversarial training framework tailored to increase the robustness of deep hedging strategies. Our approach extends pointwise adversarial attacks to the distributional setting and introduces a computationally tractable reformulation of the adversarial optimization problem over a Wasserstein ball. This enables the efficient training of hedging strategies that are resilient to distributional perturbations. Through extensive numerical experiments, we show that adversarially trained deep hedging strategies consistently outperform their classical counterparts in terms of out-of-sample performance and resilience to model misspecification. Our findings establish a practical and effective framework for robust deep hedging under realistic market uncertainties.

2 - Diffusion Factor Models: Generating High-Dimensional Returns with Factor Structures

Yumin Xu, Peking University, Beijing, China, People's Republic of

We introduce a diffusion factor model that integrates latent factor structures into generative diffusion processes to address high dimensionality and data scarcity in financial simulation. By decomposing the score function via time-varying orthogonal projections, we design neural architectures that exploit the low-dimensional factor structure of asset returns. Our analysis provides statistical guarantees depending only on the intrinsic factor dimension k , not the number of assets d , overcoming classical dimension-dependent limits. Numerical studies show superior subspace recovery in small data regimes, and empirical tests highlight economic value in portfolio construction. This work offers the first theoretical integration of factor models with diffusion processes, establishing a principled framework for scalable financial scenario generation.

3 - Arbitrage on Decentralized Exchanges

Yutian Zhou, The Chinese University of Hong Kong, Hong Kong, Hong Kong

Decentralized exchanges (DEXs) are alternative venues to centralized exchanges (CEXs) for trading cryptocurrencies and have become increasingly popular. An arbitrage opportunity arises when the exchange rate of two cryptocurrencies in a DEX differs from that in a CEX. Arbitrageurs can then trade on the DEX and CEX to make a profit. Trading on the DEX incurs a gas fee, which determines the priority of the trade being executed. We study a gas-fee competition game between two arbitrageurs who maximize their expected profit from trading. We derive the unique symmetric mixed Nash equilibrium and find that (i) the arbitrageurs may choose not to trade when the arbitrage opportunity and liquidity is small; (ii) the probability of the arbitrageurs choosing a higher gas fee is lower; (iii) the arbitrageurs pay a higher gas fee and trade more when the arbitrage opportunity becomes larger and when liquidity becomes higher; (iv) the arbitrageurs' expected profit could increase with arbitrage opportunity and liquidity. The above findings are consistent with our empirical study.

4 - The Paradox of Just-in-Time Liquidity in Decentralized Exchanges

Brian Zhu, Columbia University, New York, NY, United States

Passive liquidity providers (PLPs) in automated market makers face inevitable rent extraction from arbitrageurs, but remain profitable when trading fees outweigh positional losses. We show that just-in-time (JIT) liquidity enhances PLP profitability if order flow scales sufficiently with market depth, but otherwise displaces PLPs. To prevent this displacement, we propose a two-tiered fee structure that transfers a portion of JIT LP fee revenue to PLPs. Empirically, we find that JIT LPs are informed, and leveraging the launch of MEV-Boost as a shock to the block production process, we demonstrate that JIT liquidity complements, rather than crowds out, PLPs in practice.

MC05

Building A Level 3 A305

Office of Naval Research Program Review 2

Invited Session

ONR

Chair: David Phillips, Office of Naval Research, Arlington, VA, United States

1 - Global Guarantees for Robust Matrix Factorization: Promises of Nonconvex Formulations

Salar Fattahi, University of Michigan, Ann Arbor, MI, United States

Low-rank matrix factorization is inherently a nonconvex optimization problem. The classical approach to solving such problems involves convex relaxation techniques, often enhanced with various regularization terms. While theoretically appealing, these convex methods are rarely used in practice due to their high computational cost. Instead, practitioners typically tackle the original nonconvex formulation using simple local search algorithms, such as (sub-)gradient methods. In this talk, we explore promises of nonconvex formulations, with a focus on their optimization landscape, the design of provably convergent and efficient algorithms, and their robustness to noise and model misspecification.

2 - Semidefinite Relaxations for Dynamic Combinatorial Optimization

Alejandro Toriello, Georgia Tech ISyE, Atlanta, GA, United States, Diego Cifuentes

Review for the project "Semidefinite Relaxations for Dynamic Combinatorial Optimization", sponsored by the Office of Naval Research.

3 - Boosting Convergence and Performance of Minimax Algorithms Through Acceleration and Stochastic Methods

Quoc Tran-Dinh, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

Minimax optimization plays a pivotal role in various areas, including robust optimization, game theory, and machine learning. In recent years, numerous high-impact applications — such as adversarial training, distributionally robust optimization, reinforcement learning, and online learning — have been formulated as minimax problems, further highlighting the appeal of this research area. Our ONR project (N00014-23-1-2588) was established to investigate these problems through the lens of the optimality tracking principle, in conjunction with contemporary optimization techniques. In this talk, the PI will present recent results from this ONR project. We begin by recasting the optimality condition of minimax problems as a generalized equation or root-finding problem involving structured operators. Building on this formulation, we introduce several new algorithms that leverage both modern optimization strategies and innovative ideas such as acceleration, randomization, and stochastic techniques. We also provide a novel convergence analysis based on Lyapunov functions, which leads to new convergence guarantees that extend beyond those found in current literature. Finally, we demonstrate practical applications in statistical learning and federated learning and outline future research tasks and potential applications stemming from the results of this project.

4 - Randomized Optimization Algorithms for Large-Scale Data Systems

Clifford Stein, Columbia University, New York, NY, United States

First we address the fully-dynamic carpooling problem with recourse: Edges arrive and depart online from a graph G nodes according to an adaptive adversary. Our goal is to maintain an orientation H of G that keeps the discrepancy, defined as

$$\max_v |\text{outdegree}_H(v) - \text{indegree}_H(v)|,$$

small at all times. We present a simple algorithm and analysis for this problem with recourse based on cycles, and a reduction to low girth graphs.

Next, we address point-to-point packet routing in undirected networks, which is the most important communication primitive in most networks. The main result proves the existence of routing tables that guarantee a polylog-competitive completion-time series deterministically. In any undirected network it is possible to give each node simple stateless deterministic local forwarding rules, such that, any adversarially chosen set of packets are delivered as fast as possible, up to polylog factors. All previous routing strategies crucially required randomization for both route selection and packet scheduling. The core technical contribution of this work is a new local packet scheduling result of independent interest.

MC06

Building A Level 3 A306

E-Markets

Invited Session

eBusiness

Chair: Shengjun Mao, The University of Hong Kong, Hong Kong, Hong Kong

1 - AI as A Framing Tool: Investigating the Influence of Generative Summary on Review Generation Through A Natural Experiment

Rui QIN, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Zhaoyang YU, Katherine Feng

The integration of generative AI into digital platforms has introduced AI-generated summary (AIGS) of online reviews as a novel interface for presenting consumer opinions, potentially altering how consumers process information and articulate their own reviews. This paper empirically investigates the impact of AIGS on consumers' cognitive processing and review generation, with a particular focus on the multidimensional cue the summary conveys. Leveraging a cross-platform difference-in-differences (DID) model, we compare consumer review generation on Tripadvisor, where AIGS was introduced, with Booking, which did not adopt this feature. Our analysis shows that AIGS significantly influences both the numerical ratings and the textual content of consumer reviews. Following the introduction of AIGS, consumers tend to provide less extreme ratings and write reviews using simpler lexicon and syntax. Mediation analysis suggests that the multidimensional nature of AIGS broadens consumers' evaluative scope during cognitive processing, which in turn shapes how they formulate and express their opinions. Our study contributes to a growing literature on human-AI interaction by revealing how AIGS can act as a cognitive framing tool in digital content creation. The findings offer actionable insights for platform designers seeking to guide user-generated content and cultivate more balanced and informative online reviews through AI-driven interventions.

2 - Gov-Rag: a Retrieval-Augmented Generation Framework for Enhancing E-Government Services

Miao Yu, The University of Hong Kong, Hong Kong, Hong Kong, Hailiang Chen

The rapid evolution of Large Language Models (LLMs) has opened new opportunities for applying AI in high-stakes domains. However, deploying LLMs involves challenges like outdated training data, limited explainability, and reliability issues due to hallucinations. This study introduces Gov-RAG, a novel Retrieval-Augmented Generation framework, specifically tailored to address the unique demands of e-government applications. Grounded in trust in automation theory and domain-specific practical adaptation needs, the Gov-RAG framework achieves five key goals: (1) In-depth domain knowledge, (2) Factual accuracy in sensitive domains, (3) Human-centered explainability, (4) Low-cost real-time updates, and (5) Hallucination reduction. The framework incorporates a real-time updated comprehensive dataset, aggregating national laws, regulations, and government policies. To enhance retrieval precision, Gov-RAG combines dense and sparse vector retrieval with inverted index full-text retrieval and applies domain-specific hierarchical slicing. We also develop a fact-based evaluation framework to measure hallucinations across different LLM architectures, an aspect that was not properly measured in prior work. Rigorous computational experiments show that while fine-tuning improves stylistic alignment with standard answers, it also increases hallucination. Gov-RAG can improve factual consistency and reduce contradictions, consistently outperforming base LLMs and alternative frameworks in governmental applications. Furthermore, RAG with fine-tuned LLMs exhibits degraded performance in answering questions that require historical knowledge. Gov-RAG exhibits balanced capabilities across different test datasets. Online experiments with real citizens further validate the feasibility and impact of Gov-RAG on user performance, user perceptions, and user behaviors. Results indicate significant improvements in task performance, user-perceived explainability, and efficiency during engagement compared to base LLMs.

3 - The Economics of on-Premise LLMs

Jiaqi Shi, University at Buffalo, Buffalo, NY, United States, Zhoupeng Zhang, Shaojie Tang

Large Language Models (LLMs) like ChatGPT and DeepSeek have surged in capability and popularity, enabling diverse applications from automated customer service to content generation. LLM users now face a critical choice between utilizing cloud services and running on-premise models. This decision carries significant implications for cost, data privacy, and performance. Simultaneously, LLM service providers face the challenge of determining if and when to allow users to deploy models locally. Our study analyzes how LLM providers' choices regarding service models impact their revenue and the corresponding consumer behavior.

4 - Voice Versus Text? An Empirical Analysis on the Impact of Physicians' Voice Messages in Online Medical Consultations

Shengjun Mao, The University of Hong Kong, Hong Kong, Hong Kong

While Online Medical Consultation (OMC) platforms offer patients convenient and accessible healthcare services, challenges remain in ensuring effective communication between physicians and patients. Several OMC platforms provide physicians with a voice reply function, enabling them to send voice messages during consultations. This study aims to examine the effectiveness of such an IT-enabled function. We integrate media synchronicity theory (MST) and social support theory (SST) to develop a conceptual framework, highlighting how different media with various capabilities influence OMC performance by enabling physicians to convey specific components of social support. Specifically, we analyze data from a leading OMC platform in China to assess the impact of physicians' adoption of voice messages on OMC outcomes. The empirical results show that when physicians use voice messages to respond to patients, patients are more likely to report higher satisfaction and perceive the consultation as more helpful. To explore the underlying mechanisms, we develop advanced deep learning models to extract key factors of emotional support and informational support within OMC interactions. By conducting moderation analyses, we find that voice messages are particularly effective in conveying emotional support components like empathy and affirmation, as well as concrete informational support, which enhances the effects of voice-message use on OMC performance. Our findings suggest that healthcare stakeholders can strategically leverage the voice message feature to improve online consultation outcomes.

MC07

Building A Level 3 A307

Iterative Optimization Approaches

Job Market Showcase

Job Market Showcase

Chair: Trang Tran, Lehigh University, Bethlehem, PA, 18015, United States

1 - Machine Unlearning: is One Newton Step Enough?

Haolin Zou, Columbia University, New York, NY, United States, Arnab Auddy, YongChan Kwon, Kamiar Rahnema Rad, Arian Maleki

Machine unlearning focuses on the computationally efficient removal of specific training data from trained models, ensuring that the influence of forgotten data is effectively eliminated without the need for full retraining. Despite advances in low-dimensional settings, where the number of parameters p is much smaller than the sample size n , extending similar theoretical guarantees to high-dimensional regimes remains challenging.

We study an unlearning algorithm that starts from the original model parameters and performs a theory-guided sequence of Newton steps $(T \in \{1, 2\})$. After this update, carefully scaled isotropic Laplacian noise is added to the estimate to ensure that any (potential) residual influence of forget data is completely removed.

We show that when both $(n, p \rightarrow \infty)$ with a fixed ratio (n/p) , significant theoretical and computational obstacles arise due to the interplay between the complexity of the model and the finite signal-to-noise ratio. Finally, we show that, unlike in low-dimensional settings, a single Newton step is insufficient for effective unlearning in high-dimensional problems---however, two steps are enough to achieve the desired certifiability. We provide numerical experiments to support the theoretical claims of the paper.

2 - Improved Complexity for Nonconvex Optimization: Two-Level Online Learning with Quasi-Newton Methods

Ruichen Jiang, University of Texas at Austin, Austin, TX, United States, Aryan Mokhtari, Francisco Patitucci

A fundamental problem in optimization is finding an ϵ -first-order stationary point of a smooth nonconvex function using only gradient information. The best-known gradient complexity for this task, assuming both the gradient and Hessian of the objective function are Lipschitz continuous, is $O(\epsilon^{-1.75})$. In this talk, I will present a new method with a gradient complexity of $O(d^{0.25}\epsilon^{-1.625})$, where d is the problem dimension, yielding an improvement when $d = O(\epsilon^{-0.5})$.

Our key idea is to design a quasi-Newton method that operates through two levels of online learning. Specifically, we first reformulate the task of finding a stationary point for a nonconvex problem as minimizing the regret in an online convex optimization problem, where the loss is determined by the gradient of the objective function. Then, we introduce a novel optimistic quasi-Newton method to solve this online learning problem, with the Hessian approximation update itself framed as an online learning problem in the space of matrices. Beyond improving the known complexity bounds, our result provides the first guarantee that quasi-Newton methods can outperform gradient descent-type methods in nonconvex optimization.

3 - The Power of Adaptivity in SGD

Matthew Faw, Georgia Institute of Technology, Atlanta, GA, United States

We study convergence rates of AdaGrad-Norm as an exemplar of adaptive stochastic gradient descent (SGD) methods, which adjust their step sizes based on observed stochastic gradients, for minimizing non-convex, smooth objectives. Despite their popularity, the analysis of adaptive SGD lags behind that of non-adaptive methods in this setting. Specifically, all prior studies rely on some subset of the following assumptions which are unnecessary in the analysis of fixed step-size SGD: (i) uniformly bounded gradient norms, (ii) uniformly bounded stochastic gradient variance (or even noise support), (iii) conditional independence between the step size and stochastic gradient. In this work, we show that AdaGrad-Norm exhibits an order optimal (up to polylogarithmic factors) convergence rate, scaling inversely with the square-root of T after T iterations, under the same standard assumptions as a well-tuned fixed step-size SGD (unbounded gradient norms and affine noise variance scaling). We also go beyond the standard smoothness assumption, proving similar results for (L_0, L_1) -smooth (non-convex) functions (where the norm of the Hessian can scale affinely with the norm of the gradient, e.g., the exponential function). This is the first rate of convergence for any algorithm in this smoothness and noise regime, even for the special case of uniformly-bounded variance.

4 - Stochastic Gradient Methods Without Replacement in Finite-Sum Optimization

Trang Tran, Lehigh University, Bethlehem, PA, United States, Lam Nguyen, Katya Scheinberg

We consider the finite-sum optimization problem, which is important in machine learning, where the objective function is an average of loss functions over many data samples. Stochastic gradient descent and its momentum variants are effective and scalable for this problem. While the theoretical analysis for these methods often relies on sampling with replacement, most practical machine learning libraries use sampling without replacement (shuffling sampling) for the update sequence.

In this talk, we first briefly discuss some theoretical advantages of shuffling sampling in stochastic gradient methods. Then, we introduce Nesterov Accelerated Shuffling Gradient (NASG) algorithm for convex finite-sum minimization, which integrates Nesterov's acceleration momentum with shuffling sampling. This algorithm has a convergence rate of $O(1/T)$ where T is the number of data passes, which is better than any other shuffling gradient methods in a general convex setting. Finally, we discuss a theoretical connection with multi-objective optimization.

MC08

Building A Level 3 A310

Technology, Platforms and Society

Invited Session

Information Systems

Chair: Helen Zeng, UNIVERSITY OF CALIFORNIA - DAVIS, 540 Alumni Ln, Davis, CA, 95616-5270, United States

1 - From Views to Value: a Multimodal Analysis of Video Engagement and Learning Outcomes in Online Education

Yi-Yun Chu, Carnegie Mellon University, Pittsburgh, PA, United States

As digital learning platforms increasingly become the primary interface between learners and educational content, understanding how platform design shapes engagement and performance is critical. This study investigates how video structure, cognitive load, instructor characteristics, and learner behavior interact to influence engagement and academic outcomes in online learning environments. Using a large-scale, cross-subject dataset, we examine the nuanced relationship between cognitive demand and learning efficacy, identifying an optimal challenge zone where moderate cognitive load maximizes video completion and performance. To support this analysis, we construct a multimodal cognitive load measurement framework that captures moment-to-moment learner difficulty using interaction patterns and video structural features. We find that video structure, ranging from introductory motivational segments to core concept delivery and advanced, example-driven dense presentations, plays a key role in mediating these effects. Moreover, instructor attributes and learner replay behaviors show dynamic patterns of engagement, highlighting the potential for real-time, AI-driven interventions to support adaptive scaffolding. Overall, our findings offer actionable insights for platform designers and educators, advocating for dynamic, modular video content and behavior-aware personalization to enhance learning outcomes at scale.

2 - Quantifying the Adoption of LLM-Assisted Writing in Society: Evidence from Science, Business, and Public Discourse**Hancheng Cao, Emory University, Atlanta, GA, United States**

Generative AI tools like ChatGPT are reshaping how people write across many parts of society, yet their real-world adoption has been difficult to measure. In this talk, I present a statistical framework for estimating the presence of LLM-assisted writing across large-scale corpora. Applying this approach to diverse datasets—including scientific peer reviews, academic manuscripts, job posts, consumer complaints, and international press releases—we uncover clear and accelerating patterns of LLM adoption. These trends vary across domains, timeframes, and organizational contexts, offering insight into how generative AI is being integrated into high-stakes communication.

3 - Creative Markets in the Age of Generative AI: Strategic Shifts and Labor Market Health**Eric Zhou, Boston University, Boston, MA, United States**

The rise of generative artificial intelligence (AI) has sparked significant debate about its impact on creative labor markets, echoing concerns raised nearly 200 years ago with the advent of photography. While early photographic technologies displaced traditional portraiture forms, they also spurred artistic innovation, as painters embraced abstraction and new approaches to human expression. Today, generative AI poses a similar challenge but at an unprecedented scale, raising concerns about labor displacement, intellectual property, and the future of human creativity. This paper investigates how creators respond to competitive pressures and ethical concerns associated with generative AI by adopting, protesting, or opting out of AI training systems and the long-term implications of these strategic shifts on the health and composition of creative markets. We conduct artist- and artifact-level analyses using large-scale data from a creative labor market platform to understand strategic responses to both the introduction of AI and protections offered by the platform against AI. We investigate to what extent these strategic shifts impact the overall artist and artwork composition within the broader creative labor market and its implications on the long-term health of such markets.

4 - AI Regulation in the US: A Descriptive Analysis**Hemant Bhargava, University of California, Davis, CA, United States**

This work will be discussing trends in AI regulation among the 50 state legislatures in the US, including key topics, factors that affect success (enactment) vs failure, overlap and sequencing across states, and intensity of regulation vs various state-specific factors.

MC09

Building A Level 3 A311

Humans and the Digital World

Job Market Showcase

Job Market Showcase

Chair: Grace Babalola, Binghamton University, Binghamton, NY, United States

1 - Strategic Timing in Crowdfunding: Effect of Competition on Project Launch Decisions**Shaolin Pu, The University of Kansas, School of Business, Lawrence, KS, United States, Karthik Srinivasan, Ben Sherwood, Arvind Tripathi**

Launching a reward-based crowdfunding (RBC) campaign mirrors the complexity of starting a small venture, requiring more than a compelling idea—it demands strategic timing amid competition. While prior research has focused on storytelling and project presentation, little is known about how competitive pressure shapes creators' launch decisions. This study addresses that gap by examining whether the presence of mega (high-performing) and ordinary (newly launched) competitors influences creators' timing strategies and campaign outcomes. Using data from 184,537 Kickstarter projects (2014–2022), we analyze creators' day-level launch behavior and its effect on pledged amounts. Our findings show that creators strategically adjust their launch timing not only to benefit from the positive network effects of high-profile competitors, but also to avoid periods of intense competition that could overshadow their campaigns, and that such avoidance positively impacts funding performance. These insights highlight the importance of competition-aware timing strategies and offer practical implications for campaign planning and platform management in RBC ecosystems.

2 - Creditability and Risk of Explanations for AI Model's Predictions**Amir Asrzd, University of Massachusetts Lowell, Lowell, MA, United States**

This talk presents a framework for generating credible and risk-sensitive counterfactual explanations for machine learning (ML) model predictions. While ML and AI models often achieve high predictive accuracy, their black-box nature makes them difficult to interpret, posing significant challenges for adoption in high-stakes domains. Explainable AI (XAI) aims to address this issue by producing explanations that are meaningful to users. However, most existing XAI methods assume that model predictions are correct, leading to potentially misleading or incomplete explanations, especially when predictions are uncertain or unreliable.

This research tackles two fundamental problems in XAI: the credibility of uncertain predictions and the risk of misleading explanations. We propose new credibility measures to detect and flag unreliable predictions, and develop a novel method for generating two types of counterfactual explanations: robust counterfactuals that reduce the risk of weak or unstable explanations, and vigilant counterfactuals that sensitively detect harmful or undesirable changes.

The method is validated empirically using real-world data, demonstrating its effectiveness in improving the trustworthiness of AI-assisted decision-making. This work contributes to the development of more transparent and responsible AI systems, particularly in environments where prediction errors can have serious consequences.

3 - Human-Centered AI/ML Systems: Bridging Research and Industry Applications in Healthcare and Technology**Grace Babalola, Binghamton University, Binghamton, NY, United States**

My work showcases a comprehensive research program that integrates human factors engineering and data science with advanced AI/ML systems to optimize decision-making in complex, high-stakes environments. My work demonstrates how human-centered design principles can enhance AI system effectiveness across healthcare, web security, and e-commerce domains.

At Microsoft, I developed LLM-powered knowledge base chatbots using Azure OpenAI and RAG architectures for performance improvements through Teams integration and automated data processing. I also created heuristic algorithms for phishing detection that improved ML model performance through automated ground truth data generation.

My Amazon research focused on augmented analytics for natural language processing, where I built ensemble ML systems and NER pipelines using spaCy to streamline developer workflows and extract business insights from device hardware/software datasets.

At Excellus Blue Cross Blue Shield, I developed predictive models for chronic condition risk assessment and member sentiment analysis, optimizing KPIs and improving service delivery for 5,000+ members through interactive R Shiny applications.

My doctoral dissertation research addresses critical gaps in healthcare decision-making by developing AI-powered applications that reduce nursing cognitive load while maintaining human-centered design principles. Through comprehensive user research across 8 U.S. states, I identified workflow challenges and designed evidence-based decision support systems that facilitate peer collaboration and accelerate clinical decision-making.

This integrated approach demonstrates how rigorous human factors research can inform practical AI implementations, creating systems that augment rather than replace human expertise across diverse industry applications.

MC10

Building A Level 3 A312

Design and Analysis of Experiments

Job Market Showcase

Job Market Showcase

Chair: Mehrdad Ghadiri, MIT, Cambridge, United States

1 - Novel Experimental Design and Analysis Method for Solver Selection, Blockchain Design and Educational Games

Ramiro Rodriguez Buno, The Ohio State University, Columbus, OH, United States, Theodore Allen

We developed an experimental method to perform hyper-parameter tuning for new optimization problems with shared, nested and random factors with a large number of runs. We aim to optimize solver engines with applications in airline operations and compare the performance of the solvers. Initially, the experiments were based on standard library MILP problems, later, we developed our methodology for specific problems in the airline industry, to contrast the difference in results.

For a second problem, we create a design and analysis method to tune private blockchains models to different environments for industrial applications and manufacturing. We started comparing different blockchain simulators, protocols (block, transactions, nodes parameters) and analyze the main relevant outputs (transaction throughput, blockchain ledger, block mined, stale blocks, rewards gained by each miner) to select the best parameter set configuration. We study the robustness of the protocol and implications for cybersecurity.

In a third matter, we develop game-based methods for teaching experimentation and optimization in manufacturing and systems operations management. The idea is to maximize the benefits of a good DOE planning and strategy selection, comparing to the experimentation information improvements if the strategy selected is doing a lot of trial-and-errors or One-Factor-At-a-Time (OFAT).

2 - Personalized Policy Learning Through Discrete Experimentation: Theory and Empirical Evidence

Zhiqi Zhang, Washington University in St. Louis, St. Louis, MO, United States, Zhiyu Zeng, Ruohan Zhan, Dennis Zhang

Randomized Controlled Trials (RCTs), or A/B testing, have become the gold standard for optimizing various operational policies on online platforms. However, RCTs on these platforms typically cover a limited number of discrete treatment levels, while the platforms increasingly face complex operational challenges involving optimizing continuous variables, such as pricing and incentive programs. The current industry practice involves discretizing these continuous decision variables into several treatment levels and selecting the optimal discrete treatment level. This approach, however, often leads to suboptimal decisions as it cannot accurately extrapolate performance for untested treatment levels and fails to account for complex interactions between treatment effects and user characteristics. This study addresses these limitations by developing a theoretically solid and empirically verified framework to learn personalized continuous policies based on high-dimensional user characteristics, using observations from an RCT with only a discrete set of treatment levels. Specifically, we introduce a deep learning for policy targeting (DLPT) framework that includes both personalized policy value estimation and personalized policy learning. We prove that our policy value estimators are asymptotically unbiased and consistent, and the learned policy achieves a square root of n regret bound. We empirically validate our methods in collaboration with a leading social media platform to optimize incentive levels for content creation. Results demonstrate that our DLPT framework significantly outperforms existing benchmarks, achieving substantial improvements in both evaluating the value of policies for each user group and identifying the optimal personalized policy.

3 - Efficient Active Learning Strategies for Computer Experiments

Difan Song, Georgia Institute of Technology, Atlanta, GA, United States, Roshan Joseph

Active learning in computer experiments aims at allocating resources in an intelligent manner based on the already observed data to satisfy certain objectives such as emulating or optimizing a computationally expensive function. There are two main ingredients for active learning: an initial experimental design, which helps to approximately learn the function, and a surrogate modeling technique, which provides a prediction of the output along with its uncertainty estimates. Space-filling designs are commonly used as initial designs and Gaussian processes for surrogate modeling. This article aims at improving the active learning procedure by proposing a new type of initial design and a new correlation function for the Gaussian process. The ideas behind them are known in other fields, such as in sensitivity analysis or in kernel theory, but they never seem to have been used for active learning in computer experiments. We show that they provide substantial improvement to the state-of-the-art methods for both emulation and optimization. We support our findings through theory and simulations, and a real experiment on the vapor-phase infiltration process.

4 - Regression Adjustment for Average Treatment Effect in Finite Population Randomized Controlled Trials

Mehrdad Ghadiri, MIT, Cambridge, MA, United States, Ali Jadbabaie, Alberto Abadie, Mahyar Jafarinodeh

The design and analysis of randomized experiments are fundamental to many fields, from the physical and social sciences to industrial applications. Regression adjustment is a widely used technique for leveraging auxiliary covariate information to reduce the variance of experimental estimates. Although a large body of work in statistics and econometrics studies various regression-adjustment methods and their asymptotic properties, relatively little attention has been paid to finite-population settings with non-asymptotic accuracy guarantees.

In this talk, we consider the estimation of the average treatment effect (ATE) in finite population randomized controlled trials. We introduce unbiased ATE estimators based on both simple and complete randomization combined with regression adjustment, and we derive non-asymptotic bounds on their variance, concentration, and confidence intervals. Moreover, we discuss connections between our estimators and both classic and modern methods in experimental design and regression adjustment, and we present numerical comparisons demonstrating that our estimators outperform existing approaches in practice.

MC11

Building A Level 3 A313

High-dimensional Data Analysis I

Job Market Showcase

Job Market Showcase

Chair: Linfeng Song, North Carolina State University, Raleigh, NC, 27606, United States

1 - AI-Driven Analysis and Optimization of Systems with High-Dimensional Sensing Data

Mohammad Bisheh, Georgia Institute of Technology, Atlanta, GA, United States

Modern manufacturing and health-monitoring systems generate high-dimensional, noisy, and often incomplete sensor streams, and transforming these raw data into actionable insights requires methods that are both statistically rigorous and computationally scalable. This presentation offers an integrated statistical-learning framework, rooted in data science and advanced manufacturing, to address key challenges in additive manufacturing (AM), 3D printing, and multivariate anomaly detection across diverse industrial and biomedical settings. At its core is a robust functional tensor-decomposition (RFTD) algorithm that treats multichannel time-series data as time-

indexed tensors, enforces spatio-temporal smoothness, and iteratively separates each tensor into a low-rank “true process” component and a sparse “anomaly” component, even under asynchronous sampling and missing values, via proximal alternating minimization. Complementing RFTD, an adaptive feature-extraction pipeline combines Daubechies wavelet packets with short-time Fourier transforms to yield shift-invariant, noise-resilient descriptors; coupled with isolation forests and local outlier factors, this pipeline dramatically improves fault isolation over kernel-PCA baselines. To reduce experimental burden in metal and polymer AM, a transfer-learning scheme propagates layer-wise printability knowledge across materials, while manifold learning on electron-backscatter diffraction images clusters grains by morphology and orientation, enabling microstructure-aware monitoring with minimal labeling. Beyond manufacturing, applying RFTD-derived tensor graphs to resting-state fMRI uncovers time-varying connectivity patterns associated with ADHD and ASD, illustrating the method’s versatility for biomedical signal analysis. Collectively, these contributions deliver a scalable AI toolkit for real-time quality assurance, process optimization, and anomaly detection, advancing the reliability, efficiency, and adaptive intelligence of next-generation smart-manufacturing and health-monitoring ecosystems.

2 - Continuous-Time Modeling of Complex Operational Systems with Neural Differential Equations

YongKyung Oh, UCLA, Los Angeles, CA, United States

Many critical operational systems, from patient health trajectories in healthcare to asset movements in logistics, are inherently dynamic and generate complex, irregularly-sampled time series data. Effectively modeling these continuous-time processes to support robust and timely decision-making remains a significant challenge for conventional discrete-time analytical methods, which can struggle with asynchronous observations and underlying system dynamics. My research direction focuses on advancing and applying Neural Differential Equations (NDEs) and related continuous-time modeling frameworks to overcome these limitations. This approach allows for a more natural and powerful representation of systems that evolve continuously, directly handling irregularly-sampled data and offering the potential to uncover the governing dynamics with greater fidelity. The core of this research involves developing novel NDE-based architectures and robust learning algorithms tailored for operational data.

Beyond methodological innovation, a key objective is to ensure these advanced continuous-time models are robust, interpretable, and deployable in critical operational settings, addressing challenges like data shifts. The ultimate aim is to leverage these NDE-based modeling capabilities to significantly enhance forecasting, real-time anomaly detection, and control strategies. Applications span improving dynamic risk assessment in healthcare, optimizing traffic flow in intelligent transportation systems, and increasing the efficiency of supply chain logistics, leading to more adaptive, insightful, and effective decision support systems. This continuous-time modeling paradigm offers new avenues for tackling previously intractable challenges in the analysis and control of dynamic operational environments.

3 - An Explainable Neural Network Framework for Multi-Sensor Prognostics

Linfeng Song, North Carolina State University, Raleigh, NC, United States

Accurate prediction of failure times is critical for the effective maintenance and risk management of complex engineering systems. This study proposes an explainable neural network (XNN) framework for multi-sensor prognostics that enables interpretable failure time prediction. The framework integrates deep learning-based regression with feature attribution techniques to identify the most informative sensors contributing to the remaining useful life (RUL) estimation. Through simulated experiments, we demonstrate that the proposed model achieves competitive prediction accuracy while providing transparent insights into sensor relevance. By enhancing interpretability without sacrificing performance, this work contributes to building trustworthy, data-driven prognostic tools in industrial applications.

4 - Theoretical Analysis for Optimal Design of An Online Monitoring and Sampling Scheme Under Partial Observations

Dongmin Li, Georgia Institute of Technology, Atlanta, GA, United States, Xiaochen Xian

High-dimensional streaming data is collected in a wide range of applications, such as advanced manufacturing and environmental monitoring. However, practical constraints such as limited sensors, transmission bandwidth, and data storage capacity, restrict the ability to observe all variables of interest at each data collection time for online monitoring. Although adaptive sampling strategies have been proposed that are integrated into monitoring schemes for quick anomaly detection, the theoretical analysis of detection performance for optimal design in practical settings has not been addressed. While few studies provide upper bounds for the average detection delay, the bounds are either asymptotic or excessively large, failing to provide valuable guidance on parameter design for optimal detection performance in practical applications. To fill this gap, we propose a holistic monitoring-based probabilistic sampling framework under partial observations and theoretically analyze the average detection delay to inform optimal design. By leveraging the sampling probabilities of variables, which are explicit functions of monitoring statistics, we derive formulas for practical average detection delay as functions of design parameters and system settings, such as the number of sampled variables and total variables in the system. These theoretical results enable the determination of the number of sampled variables and parameter settings to achieve optimal detection performance while satisfying false alarm requirements and considering practical resource constraints such as sampling costs. Simulation studies across various parameters and system settings validate the effectiveness of the theoretical results, demonstrating significant practical value for optimal performance design.

MC12

Building A Level 3 A314

Supply Chain and Service Networks

Job Market Showcase

Job Market Showcase

Chair: Varun Suriyanarayana, Cornell University, Ithaca, NY, United States

1 - 3D-Printing for Supply Chain Resilience

Ziyu He, University of Southern California, Los Angeles, CA, United States, Vishal Gupta, Nick Vyas

Most strategies for mitigating supply chain disruptions require upfront, dedicated investments for each supplier, making them impractical for large, disaggregated supply chains. We study using 3D printing (3DP) as a flexible, backup resource that can support any disrupted supplier. 3DP has traditionally been rejected as a viable resilience strategy due to high per-unit production costs and limited capacity. Inspired by the “a little flexibility is enough” literature, however, we explore when and why 3DP may be a cost-effective resilience strategy and how to deploy it. We formulate an optimization model to decide which suppliers to backup with 3DP, how much 3DP capacity to procure, and how to allocate that capacity in real-time to unmet demand. The resulting mixed-binary, stochastic optimization problem is computationally challenging, even for moderate sized supply chains. Hence, we propose a novel algorithmic framework combining supermodular approximations based on Taylor series and first-order stochastic optimization to compute high-quality feasible solutions. Finally, we conduct an empirical case study based on bill-of-lading data from toy manufacturer Mattel. With cost estimates based on current technology, 3DP offer modest cost savings relative to traditional resilience strategies. However, its principal benefit is a reduction in demand shortfalls. This advantage is especially evident in larger systems with weakly correlated supplier disruptions. Resonating with the well-known “a little flexibility is enough” principle, our findings provide both analytical and empirical evidence of 3DP’s transformative potential as a strategic resilience tool for large supply chains.

2 - Designing Spare Part Supply Chains for High-Tech Products: a Life Cycle Approach Using a Bayesian Demand Model

Firat Kilci, The University of Texas at Austin, Austin, TX, United States, Genaro Gutierrez, Saurabh Bansal

Spare parts are an essential aspect of after sales service to increase product economic life, customer satisfaction, and future sales. In this study, we partner with a global computer manufacturer and examine the structure of a multi-echelon service network for spare parts in the high-tech industry over the course of their life cycle under a Bayesian demand model. High-tech components have short life cycles and are under risk of obsolescence. Overstocking these components results in high inventory holding costs in the form of depreciation and low salvage values, and understocking results in a decrease in service levels, emergency shipments costs and a possible decrease in future sales. We formulate the problem as a single-item multi-echelon Bayesian dynamic inventory problem where the aim is to decide on an inventory policy to minimize the total expected inventory management costs. We derive a near-optimal policy and provide bounds on both the policy and cost spaces. Using real data from our research partner, we use our model to derive the dynamic behavior of the structure of a multi-national supply chain network that includes 28 countries. Based on the

results, which capture the dynamic learning of demand model parameter values, we identify different dynamic behaviors corresponding to the introduction, growth, maturity and decline phase of the product life cycle. This model informs management that a near-optimal global supply chain structure is a dynamic amalgamation of country specific policies responding to their own parameters, and it helps them to implement these policies.

3 - Rethinking Fulfillment Regionalization

Nidhima Grover, Georgia Institute of Technology, Atlanta, GA, United States, Daniel Ulch, Xiaoyan Si, Yuan Li, Louis Faugère, Amitabh Sinha, Benoit Montreuil

In today's retail industry with the growing opportunity of providing delivery of goods directly to consumers, companies can store their inventory at cheaper locations away from the high demand locations of customers. This provides them with the opportunity of pooling inventory at Fulfillment Centers (FCs), storing lower levels of safety stock. Recently, retailers started decomposing their fulfillment network into smaller regions through a method called Regionalization, which provided improvements in cost and speed of fulfillment. In every region, fulfillment is prioritized from a set of FCs assigned to the region, and every FC is assigned to a single region (single mapping). In this paper, we study a variant of regionalization where we can assign FCs to more than one regions (multi-mapping). This can lead to improvements in cost and speed due to reduction in in-region distances, primarily due to the effect of inventory pooling from making an FC's inventory easily accessible to more than one region. We introduce a modeling and solution framework to design regions and assign FCs to more than one region simultaneously, using a non-linear mixed integer program. Moreover, we also introduce a novel evaluation framework where we first find the percentage of regional demand that we can fulfill from FCs assigned to a region. This helps us derive inventory allocation strategies that we further employ in a placement-fulfillment simulator, which helps us demonstrate the significance of multi-mapping compared to single mapping.

4 - New Algorithms for Joint Replenishment Incorporating Modern Considerations

Varun Suriyanarayana, Cornell University, Ithaca, NY, United States, Varun Sivasankar, Siddharth Gollapudi, David Shmoys

The joint replenishment problem (JRP) is a classical inventory management problem. We consider a natural generalization with outliers, where we are allowed to reject (that is, not service) a subset of demand points. In this paper, we are motivated by issues of fairness- if we do not serve all of the demands, we wish to “spread out the pain” in a balanced way among customers, communities, or any specified market segmentation. One approach is to constrain the rejections allowed, and to have separate bounds for each given customer. In our most general setting, we consider a set of C features, where each demand point has an associated rejection cost for each feature, and we have a given bound on the allowed rejection cost incurred in total for each feature.

We give the first constant approximation algorithms for the fairness-constrained JRP with a constant number of features; specifically, we give a 2.86-approximation algorithm in this case. Even for the special case in which we bound the total (weighted) number of outliers, this performance guarantee improves upon bounds previously known for this case. Our approach is an LP-based algorithm that splits the instance into two subinstances. One is solved by a novel iterative rounding approach and the other by pipage-based rounding.

MC13

Building A Level 3 A315

Dynamic Programming and Reinforcement Learning

Job Market Showcase

Job Market Showcase

Chair: Imon Banerjee, Northwestern University, West Lafayette, IN, United States

1 - Multi-Agent Queueing Network: Coordination with Learning

Zhaoxuan Wei, Institute of Operations Research and Analytics, NUS, Singapore, Singapore, Andrew Lim, Hanqin Zhang

We study a coordination problem in a decentralized Jackson network with an admission decision maker and multiple service providers. The admission decision maker controls the exogenous job arrivals by dynamically setting the admission fees while the service providers decide their job-processing rates. The system is operated in a decentralized way: all agents optimize their own objective subject to their local models which may mis-specify the system components controlled by other agents. We show that the decentralized queueing network can be coordinated by the transfer mechanism solving a fixed-point equation. Under the optimal transfer mechanism, the decentralized system welfare achieves the optimal centralized system welfare. The algorithm for computing the optimal mechanism and its convergence rate are established. Furthermore, we propose a reinforcement learning (RL) algorithm for the case where each decision maker interacts with its potentially biased local generative model. Our RL algorithm is shown to converge to the optimal transfer mechanism and the sample complexity for the RL algorithm is given.

2 - Mean-Variance Portfolio Selection by Continuous-TIME Reinforcement Learning: Algorithms, Regret Analysis, and Empirical Study

Yilie Huang, Columbia University, New York, NY, United States, Yanwei Jia, Xunyu Zhou

We study continuous-time mean-variance portfolio selection in markets where stock prices are diffusion processes driven by observable factors that are also diffusion processes yet the coefficients of these processes are unknown. Based on the recently developed reinforcement learning (RL) theory for diffusion processes, we present a general data-driven RL algorithm that learns the pre-committed investment strategy directly without attempting to learn or estimate the market coefficients. For multi-stock Black-Scholes markets without factors, we further devise a baseline algorithm and prove its performance guarantee by deriving a sublinear regret bound in terms of Sharpe ratio. For performance enhancement and practical implementation, we modify the baseline algorithm and carry out an extensive empirical study to compare their performance, in terms of a host of common metrics, with a large number of widely used portfolio allocation strategies on S&P 500 constituents. The results demonstrate that the proposed continuous-time RL strategy is consistently among the best especially in a volatile bear market, and decisively outperform the model-based continuous-time counterparts by significant margins.

3 - Cost-Aware Bayesian Optimization with Adaptive Stopping via Gittins Indices

Qian Xie, Cornell University, Ithaca, NY, United States, Linda Cai, Theodore Brown, Raul Astudillo, Peter Frazier, Alexander Terenin, Ziv Scully

Hyperparameter optimization is crucial in real-world applications ranging from machine learning model training to adaptive experimentation in science, engineering, and e-commerce. These scenarios are often modeled as black-box functions, which take hyperparameters as inputs and output performance metrics. Bayesian optimization is a powerful framework for efficiently optimizing such black-box functions, especially when evaluations are expensive. However, practical considerations such as varying evaluation costs, adaptive stopping time, and observable multi-stage feedback remain under-explored in this framework.

This talk is based on two of my recent works leveraging Gittins indices, which are inherently cost-aware, stopping-aware, and feedback-aware, through connections to Pandora's Box problems and Markovian/Bayesian bandits, where the Gittins policy is Bayesian-optimal. This perspective leads to a novel acquisition function together with a tuning-free stopping rule for Bayesian optimization, supported by theoretical guarantees and validated by competitive empirical performance.

I will also briefly discuss ongoing work on large language model (LLM)-driven neural architecture search for reinforcement learning (RL) state encoding, and applications to mixed-autonomy traffic control, illustrating how my research extends beyond Bayesian optimization to tackle broader challenge in mixed-type hyperparameter and architecture optimization.

4 - Statistics on Controlled Markov Chains

Imon Banerjee, Northwestern University, West Lafayette, IN, United States

MC14

Building A Level 3 A316

E-commerce and Digital Marketing

Contributed Session

Contributed

Chair: Chu Chu, Huazhong University of Science and Technology, Wuhan, China, People's Republic of

1 - Unveiling NFT Price Dynamics: the Interplay of Traits, Market Behaviors, and Strategic Manipulations

Micheli Liu, Emory University, Atlanta, GA, United States, Rajiv Garg

The rapid rise of Non-Fungible Tokens (NFTs) has introduced both transformative opportunities and challenges in digital markets, particularly in terms of valuation mechanisms and trading behaviors. This study investigates the determinants of NFT valuations by interacting intrinsic NFT traits, and buyer and seller behaviors, using comprehensive data from the Bored Ape Yacht Club (BAYC) collection. Employing a combination of ordinary least squares, random forest model, fixed-effects linear regression, and cross-sectional elastic-net methodologies, we find that intrinsic visual traits significantly influence initial valuations, maintaining their relevance throughout an NFT's lifecycle. Furthermore, interaction effects between these intrinsic traits and transactional behaviors highlight that buyer and seller reputation, as well as transaction history, significantly affect NFT prices. Notably, strategic price manipulations, facilitated by prominent market actors ("whales") and complex buyer-seller interactions, emerge as critical factors influencing market dynamics. These insights enhance the understanding of NFT markets and provide a foundation for regulatory frameworks.

2 - How Do Chatbots Shape Donors' Decisions in a Fundraising Project?

Maryam Noorymotlagh, University of Texas at Arlington, Arlington, TX, United States, Samaneh Samiee, Jie Zhang

This research examines the pivotal role of chatbots in influencing donors' behavior and decisions within a nonprofit fundraising organization. The novelty lies in the fact that, unlike previous studies comparing chatbots' performance with that of human agents, this study compares the performance of different chatbots. This study employs a field experiment as its methodology.

The findings reveal that empathic chatbots, which employ cognitive and affective empathy in conversation, significantly increase donors' engagement in fundraising campaigns. Specifically, chatbots that displayed a picture of a little girl during a fundraising event while conversing with donors enhanced both the number of contributions and donor engagement.

Moreover, this research highlights the potential of designing a chatbot as a virtual assistant for a nonprofit organization. It not only reduces labor costs but also fosters sustainable donor engagement.

3 - From Waiting to Winning: Impacts of Virtual Queue Services on Local Businesses

Hyeonju Ju, Korea University, Seoul, Korea, Republic of, Eric Kwon, Jin Soo Han, Gunwoong Lee

Managing wait times is a critical challenge in service industries, directly influencing customer satisfaction and revenue. Virtual queue services have emerged as a strategic solution, allowing businesses to manage wait times efficiently through online platforms or mobile applications. Despite their growing adoption, the broader impact of these services on individual customers and local businesses remains underexplored. This study addresses this gap by examining the spillover effects of virtual queue services on customer engagement and local business performance. Utilizing advanced econometric models, the research examines the influence of virtual queue services on customer decisions to engage with local businesses. Our preliminary results empirically substantiate the pivotal role of virtual queue services in vitalizing local businesses. Our findings contribute to a deeper understanding of the interplay between technology, wait time management, and local economies, providing valuable insights for local businesses, policymakers seeking to foster a more interconnected and vibrant local business environment, and business practitioners striving to optimize efficiency through digital solutions.

4 - Dynamic and Static Item Allocation for the in-Store Front Warehouses

Chu Chu, Huazhong University of Science and Technology, Wuhan, China, People's Republic of, Shiming Deng, Wen Luo

In response to furious competition from emerging online retailers, traditional supermarkets begin leveraging their brick-and-mortar store chains by setting up small in-store front warehouses to facilitate order picking processes for online orders. In this paper, we investigate how to select products and allocate space for the in-store front warehouse under dynamic and static strategies to minimize the total order-picking workload. For the dynamic allocation problem, we formulate it as a mixed-integer programming (MIP) model, which can be decomposed into a master multi-choice knapsack problem (MCKP) and an order picking problem (OPP) as sub-problems. An exact algorithm and a heuristic that is 1/2-bounded under certain routing policies are proposed. For the static allocation problem, we construct a stochastic gradient estimator of the expected travel distance by infinitesimal perturbation analysis (IPA) and introduce a discrete Frank-Wolfe algorithm to solve it. Numerical results reveal that the logic of dynamic allocation for the in-store front warehouse is different from that for regular warehouses, and the storage location assignment for the shelf area in supermarkets greatly impact the performance of these allocation strategies. The experimental results using data from a real retailer validate the effectiveness of proposed algorithms.

MC15

Building A Level 4 A401

Industry Insights Panel - The Future of Urban Mobility in Transportation and Delivery

Invited Session

Air Transportation Section

Chair: Yuqian Dong, Walmart, Seattle, United States

1 - Urban Air, Suburban Peace: Solving the Noise Puzzle for Aerial Mobility

Laura Morejón Ramírez, Whisper Aero, Crossville, TN, United States, Mark Moore

Aerial Drone Delivery and Drone as First Responder are positioned to be significant first mover markets that experience an inflection point where scaled operations will be taking place. Already, targeted initial markets are seeing higher volume operations by providers such as Google Wing and Zipline for Walmart. While this market incubation can take place with early drones that have unrefined characteristics, scale operations will demand solutions that ensure a high-quality community integration and user experience. One of the most rampant complaints associated with existing deliveries is the relatively high noise levels from even small drones, particularly because these deliveries are happening at their residences in quiet suburban background noise environments. Scale deliveries mean frequent neighborhood incursions with flights occurring at 300-foot altitudes where the drones can be easily heard across many neighboring households and along the trip path. The noise concern is particularly concerning during the final delivery stage where high hover power in close proximities to residences increases the neighborhood noise exposure. Whisper Aero has developed a breakthrough in low noise thrust with demonstrations that have shown the ability for a 55 lb Group 2 drone to be inaudible at altitudes of 200 feet, and noise characteristics that blend into 45 dB background levels found across most suburban neighborhoods at distances of less than 100 feet. Technology characteristics and example noise results will be shared to showcase the ability of delivery drones to provide refined characteristics that can enable these markets to achieve volume deliveries that enable economies of scale, while minimizing neighborhood annoyance.

2 - Parametric Urban Air Mobility Operational Flight Trajectory Modeling and Noise Assessment

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

Urban air mobility (UAM) is an emerging transformational flight concept for future aviation. New flight operations are expected to be carried by advanced aerial vehicles, such as uncrewed aerial systems (UAS) for package delivery and electric vertical takeoff and landing (eVTOL) aircraft for air taxi services. Although UAM promises

significant benefits, it also introduces new challenges, particularly related to environmental noise. Unlike conventional aircraft, UAM vehicles often feature distributed propulsion systems and operate frequently at low altitudes, raising concerns about localized community noise impacts. The objective of this research is to develop a parametric modeling framework to simulate UAM flight operations and assess the associated community noise exposure. The framework consists of four major elements: 1) determining potential origin and destination (O/D) sites, 2) constructing O/D pairs using a flight demand model developed based on geographic and census data, 3) generating operational flight trajectories via a grid-based path planning algorithm constrained by airspace regulations, and 4) simulating noise propagation along the trajectories between operating vehicles and virtual receivers. The proposed methodology is applied to two use cases: UAS package delivery in Atlanta area and eVTOL passenger transport in the San Francisco Bay Area. Results provide insights into the spatial distribution of noise impacts and support data-driven planning for sustainable UAM deployment.

3 - The Future of Urban Mobility in Transportation and Delivery

Surnaik Srivastava, Walmart, Santa Clara, CA, United States

The drone delivery revolution isn't about flying technology—it's about reimagining how goods and services move through urban spaces. While the hardware has matured dramatically, successful deployment hinges on understanding exactly what value customers need and how to deliver it consistently. This session cuts through the hype to examine what makes drone programs succeed or fail. Drawing from real-world pilots across diverse cities and sectors, we'll explore how to identify genuine customer needs that drones can uniquely address, from emergency medical supplies to routine e-commerce fulfillment. We'll walk through the critical decisions that shape drone value propositions: which use cases to prioritize, how urban context influences service design, and what operational trade-offs determine long-term viability. The emphasis is on practical evaluation—how do you assess market readiness, navigate regulatory landscapes, and build services that scale beyond pilot programs? Rather than promoting technology for its own sake, we'll examine how to align drone capabilities with genuine customer problems, regulatory requirements, and business economics. The result is a clear methodology for assessing drone opportunities and building programs that scale. With insights drawn from omnichannel retail, this session is ideal for anyone in transportation, logistics, or strategy roles looking to stay ahead of the curve.

MC16

Building A Level 4 A402

Supply Chain Optimization at Amazon

Invited Session

The Practice Section of Informs

Chair: Garrett van Ryzin, Amazon, New York, NY, United States

1 - Vendor Collaboration Using Consensus Planning

Garrett van Ryzin, Amazon, New York, NY, United States

Consensus Planning Protocol (CPP) is a distributed optimization framework used at Amazon to modularize and scale supply chain optimization problems. In this talk, we present our ongoing work on extending CPP as a platform for collaboration with external vendors. This involves modifying the CPP subsidy payments to make them incentive compatible using mechanism design principles. We describe both concepts and implementation.

2 - Multiple Response Agents: Fast, Feasible, Approximate Primal Recovery for Dual Optimization Methods

Yicheng Bai, Amazon, New York, NY, United States, Tetiana Parshakova, Garrett van Ryzin, Stephen Boyd

We consider the problem of minimizing the sum of agent functions subject to affine coupling constraints. Dual methods are attractive for such problems because they allow the agent-level subproblems to be solved in parallel. However, achieving primal feasibility with dual methods is a challenge; it can take many iterations to find sufficiently precise prices that recover a primal feasible solution, and even with exact prices primal feasibility is not guaranteed without strict convexity. To overcome this limitation, we propose a novel primal recovery method, multiple-response agents (MRA), that is able to rapidly reduce primal infeasibility, tolerating some degree of suboptimality, and can be used with any dual algorithm.

3 - Multi-Objective Optimization in Assortment Planning for Speed Deliveries

Suleyman Karabuk, Amazon, Bellevue, WA, United States, Ozan Candogan, Jonathan Jonker, Nanjing Jian

Assortment planning for Amazon's Sub-Same Day (SSD) program involves balancing multiple competing objectives, including selection breadth, depth, Contribution Profits (CP), Customer Instock Value (CIV), Demand Penetration, and Glance Value In-Stock Speed (GVIS), etc. Traditional approaches maximize weighted sums of these objectives, exploring trade-offs by varying weights across grids to characterize the Pareto frontier. However, this method becomes computationally expensive, especially when dealing with a large number of objectives. To address this challenge, we present a novel performance space characterization technique inspired by cutting-plane methods that iteratively constructs polyhedral inner and outer approximations of achievable performance. This approach efficiently maps the performance space—the set of all attainable objective values under various policies—thereby fully characterizing the Pareto frontier. This approach has been successfully applied to Amazon Search ranking. In this paper, we extend this methodology to Amazon's SSD assortment planning.

MC17

Building A Level 4 A403

Use Cases in Data Driven Decision Making

Invited Session

Practice Track

Chair: Dayana Cope, Delta Air lines, 200 Mimosa Dr, Decatur, GA, 30030, United States

1 - Welcome Aboard Southwest Airlines Flight 737: Optimizing Pilot Training

Whitney Conner, Southwest Airlines, Louisville, KY, United States, Megan Kendrick

Sit back and relax as we embark on an adventure through Pilot training at Southwest Airlines. We're going to have an easy take-off where we learn about a Pilot's journey from New Hire to Captain. At cruising altitude, we'll give you the 30,000 foot view of 5 MILP Optimizers that we use to schedule training events and assign trainees and facilitators. As we reach our final destination, we'll dive into the details of the model that schedules hundreds of training events each month, efficiently using limited training resources and balancing other operational priorities.

2 - How Delta Air Lines Approaches New Technology: Strategies for a Successful Tech Journey

Amin Lalani, Delta Air Lines, Atlanta, GA, United States

Despite the rapid pace of innovation, customers still expect human-centered support. While self-service options provide convenience, many prefer to speak to someone to ensure their concerns are acknowledged. However, technology should not hinder the human experience; instead, it can serve as a tool to improve interactions and outcomes.

One of the digital product leaders of Airport Customer Service Tools and Technology, Amin Lalani, will share his insights on implementing new technology at Delta Air Lines. With a strong emphasis on customer focus, Amin and his team aim to create seamless, human-first experiences while enabling the frontline to deliver operational reliability.

He will introduce frameworks for initiating transformation and leading with human-centered initiatives, covering:

- Strategies for customer education and engagement
- Emerging trends in Employee and Customer Experience technology
- Tips for enhancing collaboration and securing buy-in across teams and organizations

3 - Gain Efficiency Using Generative Ai

Indu Tiwari, Delta, Apple Valley, MN, United States, Purnajyoti Bhaumik, Pavithra Karthikeyan, Mohammed Behbooei

Delta Air Lines is harnessing the transformative power of Generative AI (GenAI) to enhance operational efficiency, empower employees, and elevate customer experiences. By integrating GenAI with traditional AI and machine learning models, Delta is creating hybrid (ML, NLP, and machine vision) solutions that are both scalable and cost-effective. These technologies are being deployed across various areas—from backend operations like crew scheduling and supply chain management to customer-facing services such as chatbots and virtual assistants that handle routine inquiries. This allows human agents to focus on more complex issues, often supported by Retrieval-Augmented Generation (RAG) systems. GenAI can also generate real-time reports, forecasts, and summaries, enabling faster, context-aware decision-making for executives and staff alike.

The company has seen measurable benefits, including a 10% decrease in customer handling time (CHT) and reduced escalations at call centers. Employees across departments are accelerating their productivity with tools like Copilot and Q for Developer. With every customer journey generating vast amounts of data, Delta emphasizes the need for standardized, enterprise-grade GenAI products governed by robust evaluation methods. These go beyond classical AI techniques to ensure quality, safety, and user-friendliness.

Recognizing the importance of responsible innovation, Delta has implemented a comprehensive governance framework that prioritizes ethical, secure, and transparent AI usage. This includes strict guidelines on data handling, human oversight, and risk management to maintain trust and regulatory compliance. Through a combination of internal development and third-party solutions, Delta is strategically leveraging GenAI to drive innovation and create meaningful value for both customers and employees.

MC18

Building A Level 4 A404

OR for Defense Planning

Invited Session

Military and Security

Chair: Isabella Sanders, Arizona State University, Newburgh, NY, United States

1 - Retrieval-Augmented Generation for Adversarial Defense Planning

Tahir Ekin, Texas State University, San Marcos, TX, United States, Laxmi Shaw, Krishna Sai Mandadapu

Capturing the uncertainty in attacker goals and strategies has long been a critical challenge in defense planning. This talk introduces a novel approach that leverages Retrieval-Augmented Generation (RAG) to simulate diverse attack scenarios and help model attacker uncertainty. By generating realistic adversarial actions and goals based on historical data and expert knowledge, our method provides a dynamic and adaptable representation of potential threats. These simulated scenarios are integrated into a probabilistic decision-making framework using Adversarial Risk Analysis, enabling defense planners to assess risks and optimize responses under uncertainty. We further discuss the computational techniques required to scale this approach, making it suitable for complex defense environments with multiple decision-makers. This approach provides a robust foundation for adaptive, data-driven defense planning in adversarial settings.

2 - The Autonomous Aircraft Search and Service Problem

John Goodwill, Air Force Institute of Technology, WPAFB, OH, United States, Matthew Robbins, Brian Lunday, Nicholas Yielding

Autonomous air vehicles (AAVs) subjected to partial observability and intermittent communications must make sequential decisions under uncertainty while performing a team search and service mission. We formulate a Markov based sequential decision process model of this autonomous aircraft search and service (A2S2) problem and solve it using a blended optimization and reinforcement learning based approach. The team of AAVs must collectively locate and service multiple, moving ground-based targets. We develop a centralized training, decentralized execution (CTDE) multi-agent reinforcement learning (MARL) algorithm that periodically conducts re-planning via a mathematical program, informing subsequent AAV decisions. MARL policies generated by our algorithm are compared to those generated by independent deep Q-networks (IDQN). Designed computational experiments investigating several problem instances show differences with respect to quality of solution, computational efficiency, and robustness over varying mission conditions. Preliminary experimentation suggests that our novel algorithm outperforms established MARL policies.

3 - Simulation Based Integrated Airline Recovery

Nirav Lad, The Air Force Institute of Technology, WPAFB, OH, United States, Lance Champagne

Disruptions occur regularly in transportation and impact power project capabilities. A disruption is an interruption in the normal course of scheduled operations. Similarities exist between military and civilian air transportation operations. Moreover, military air transportation operations are reliant on civilian airlines participating in the Civil Air Reserve Fleet (CRAF) for power projection. This research focuses on civilian airline operations due to data availability for simulation and acknowledges the similarities in military and civilian operational procedures. Airline Operation Control Centers (AOCCs) manage and mitigate disruptions by rescheduling flight operations to minimize losses and maximize passenger throughput. This multi-objective problem is known as the airline recovery problem and consists of three sub-problems addressing airplanes, crews, and passengers respectively. The aircraft recovery sub-problem is generally the first of these three sub-problems solved with the assumption that passenger impacts are accounted for and crews are available. Initial study for the airline recovery problem focused on optimization techniques to solve the aircraft recovery sub-problem within limited scenarios. Recently, machine learning approaches have shown promise at solving integrated airline recovery. Simulation-based approaches remain scarce and can accommodate reinforcement learning as well as optimization methods. This research contributes an integrated airline recovery problem simulation-based approach implementation that leverages reinforcement learning to aid decision makers in aircraft and schedule recovery. Additionally, the research explores synergy with the Digital Twin concept for air mobility operations.

4 - Weighted Rank Aggregation for Portfolio Optimization: Visualizing and Making Sense of Multiple Ranked Lists

Tyler Perini, U.S. Naval Academy, Annapolis, DC, United States

The problem of interpreting or aggregating multiple rankings is common to many real-world applications. Perhaps the simplest and most common approach is a weighted rank aggregation, wherein a (convex) weight is applied to each input ranking and then ordered. This work presents a new tool for visualizing and displaying ranking information for the weighted rank aggregation method with three input ranked lists. Traditionally, the aim of rank aggregation is to summarize the information from the input rankings and provide one final ranking that hopefully represents a more accurate or truthful result than any one input ranking. While such an aggregated ranking is, and clearly has been, useful to many applications, it also obscures information. With a weight set visualization, we show the wealth of information that is available for all alternative rank aggregations due to its nice structure. We apply weight set decomposition to the set of convex multipliers, study the properties useful for understanding this decomposition, and visualize the indifference regions. This methodology reveals information—that is otherwise collapsed by the aggregated ranking—into a useful, interpretable, and intuitive decision support tool. Included are multiple illustrative examples, along with heuristic and exact algorithms for computing the weight set decomposition.

5 - Tropical Geometry in Neural Networks: a Novel Defense Against Adversarial Attacks

Kurt Pasque, Navy, Pensacola, FL, United States

Neural networks, though powerful, are highly susceptible to adversarial attacks, which can severely impact their reliability and performance in critical applications, especially within the Department of Defense (DoD). This thesis employs Tropical Convolutional Neural Networks (TCNN), a novel technique leveraging tropical geometry to enhance neural network robustness against adversarial attacks. By reconfiguring neural networks to use a tropical embedding layer as the final layer, this

approach significantly improves resistance to attacked input data without compromising model accuracy, complexity, or training time. The research includes comprehensive experiments on image classification tasks across multiple datasets, demonstrating that TCNNs maintain higher accuracy under adversarial conditions compared to identically structured traditional models. This work not only contributes to the advancement of secure artificial intelligence (AI) applications in defense but also opens new avenues for future research in neural network robustness.

MC19

Building A Level 4 A405

Control of Stochastic Networks

Invited Session

Applied Probability Society

Chair: Nasser Barjesteh, University of Toronto, Toronto, ON, Canada

Co-Chair: Amir Alwan, University of Wisconsin-Milwaukee Lubar College of Business, Pewaukee, 53202, United States

1 - Analysis of Interactions Between Strategic Customers and Strategic Servers

Yueyang Zhong, London Business School, London, United Kingdom

Service system design is often informed by queueing theory. Traditional queueing theory assumes that customers are indefinitely patient and servers work at constant speeds. That is reasonable in computer science and manufacturing contexts. However, customers and servers in service systems are people, and, in contrast to jobs and machines, often make strategic joining and work speed decisions. We develop a game-theoretic many-server Markovian queueing model with a finite buffer in which the buffer is endogenously induced by strategic customers' joining decisions, and the work speeds emerge as the solution to a non-cooperative game among strategic servers. In an asymptotic regime in which demand becomes large, we establish necessary and sufficient conditions for the existence of individual equilibria. Based on our explicit characterization of individual equilibria and the social optimum, we also explore the behavior of customer welfare, server welfare, and social welfare at equilibrium and under the social optimum.

2 - A Whittle Index Approach to Schedule Jobs with Time-Varying Holding Cost

Zhouzi Li, Carnegie Mellon University, Pittsburgh, PA, United States, Keerthana Gurushankar, Mor Harchol-Balter, Alan Scheller-Wolf

Scheduling a stream of jobs whose holding cost changes over time is a classic and practical problem. Specifically, each job is associated with a holding cost (penalty), where a job's instantaneous holding cost is some increasing function of its class and current age (the time it has spent in the system since its arrival). The goal is to schedule the jobs to minimize the time-average total holding cost across all jobs.

The seminal paper on this problem, by Van Mieghem in 1995, introduced the generalized c-mu rule for scheduling jobs. Since then, this problem has attracted significant interest but remains challenging due to the absence of a finite-dimensional state space formulation.

Consequently, subsequent works focus on more tractable versions of this problem.

This paper returns to the original problem, deriving a heuristic that empirically improves upon the generalized c-mu rule and all existing heuristics.

Our approach is to first translate the holding cost minimization problem to a novel Restless Multi-Armed Bandit (R-MAB) problem with a finite number of arms. Based on our R-MAB, we derive a novel Whittle Index policy, which is both elegant and intuitive.

3 - Optimal Batched Scheduling of Stochastic Processing Networks Using Atomic Action Decomposition

Zhanhao Zhang, Cornell University, Ithaca, NY, United States, Jim Dai, Manxi Wu

Stochastic processing networks have broad applications in healthcare, transportation, and communication networks. Despite advancements in MDP algorithms and computational resources, optimizing control policies for these networks remains challenging due to the curse of dimensionality, as the number of feasible server assignments grows exponentially with the number of servers.

We observe that many of these systems have the reward decomposability property, where the total reward from the joint assignment of all servers can be expressed as the sum of individual server rewards. By leveraging this property, we propose the atomic action decomposition framework that reduces the original MDP, whose action size scales exponentially with the number of servers, to a smaller MDP that has a constant action size, without loss of optimality. We demonstrate that this property holds widely across various applications.

Moreover, by integrating our atomic action decomposition into a popular RL algorithm PPO, we empirically show that our approach yields superior control policies in applications such as internet switch scheduling, hospital inpatient admissions, and ride-hailing.

4 - Greedy Matching of Impatient Agents: the Role of Inventory

Angela Kohlenberg, University of Michigan, Ann Arbor, MI, United States

We study dynamic matching with heterogeneous demand and supply that arrive over time and abandon if not matched before their patience expires. A match consists of a compatible demand-supply pair, and yields a demand-dependent reward. At each moment, a controller decides whether to match and, if so, which pair to match. A fundamental trade-off underlies optimal match timing: greedy (immediate) matching ignores the benefit of holding inventory (waiting supply) for better future matches, while delayed matching risks abandonment of supply or demand. We propose the family of Restricted Greedy (RG) policies, which immediately (i.e., greedily) executes matches within an optimized subset of compatible demand-supply pairs and does not perform matches outside this subset. A knapsack-like algorithm identifies the network design: the optimal subset of the compatibility graph. We identify conditions under which RG is scale optimal: its reward loss due to abandonment is on the same order as the optimal policies for all relevant parameters (arrival rates, impatience rates, and match rewards). These conditions reveal why, in the presence of abandonment, static network design combined with greedy matching is often near-optimal. They formalize the intuition that abandonment compromises the controller's ability to fine-tune inventory levels, and therefore, system performance.

5 - Dynamic Control of Multiclass Omnichannel Hybrid Production Systems: Applications to Dynamic Pricing in Restaurants

Amir Alwan, University of Wisconsin-Milwaukee Lubar College of Business, Pewaukee, WI, United States, Nasser Barjesteh

The rise of omnichannel production systems has transformed various industries, introducing new challenges in demand and operations management. One notable development is the adoption of dynamic pricing in omnichannel restaurants. Motivated by these trends, we develop a stochastic processing network model for an omnichannel hybrid production system serving heterogeneous, price- and delay-sensitive customers. The firm offers a mix of make-to-order (MTO) and make-to-stock (MTS) goods, available through both walk-in and online channels, at multiple quote times. The objective is to maximize the long-run average expected profit through dynamic pricing, production scheduling, and order rejection decisions. MTO orders incur both earliness and tardiness costs, while MTS orders face tardiness and inventory holding costs. Walk-in customers have finite patience and may abandon the system if their waiting time is excessive. Since this problem appears analytically intractable, we consider an approximating Brownian control problem in the heavy-traffic regime. We prove that the optimal policy is a two-sided barrier policy with a state-dependent drift rate, for which we derive a closed-form solution. We then propose a joint dynamic pricing, scheduling, and rejection policy by interpreting this solution in the context of the original production system. Finally, we demonstrate the effectiveness of our proposed policy through a comprehensive simulation study and offer several managerial insights tailored to our canonical example of an omnichannel restaurant.

MC20

Building A Level 4 A406

Operations Research and Analytics at Norfolk Southern

Invited Session

Railway Applications

Chair: Yudi Pranoto, Norfolk Southern Corporation, Atlanta, GA, United States

1 - Synchronize Train Schedules for Freight Rail**Peiheng Li, Norfolk Southern Corporation, Decatur, GA, United States**

Train scheduling contributes the last fundamental decision in an operating plan for freight rail in North America following closely behind blocking and train routing. This problem has combinatorial complexity in nature as it concerns designing train schedules in terms of frequency, operating days, and timing. A scheduling plan involves synchronized train movements and shipment flows. Its evaluation often resorts to simulation. These two features make this real-life problem extremely challenging to solve. This presentation has two goals: (1) review this problem especially its relationship with these train scheduling and timetabling problems reported in literature; (2) summarize an approximation model we develop at Norfolk Southern (NS) as an endeavor on solving this problem. This approximation model is reasonable in accuracy and easy to solve. The numerical tests against the real-life operating plans have proven its effectiveness and efficiency in facilitating the transit of shipments. This approximation model has been successfully integrated with the existing scheduling decision support system at NS and has been rolled out to help design efficient schedule plans and drive productivity.

2 - Or Application at Ns Operations Center**Zahed Shahmoradi, Norfolk Southern, Atlanta, GA, United States**

Norfolk Southern aims to lead safety and efficiency in the operations. The Operation Research team is tasked to advance this goal by applying OR concepts, AI, and simulation to provide insights on optimal tonnage ratings, power assignments, and infrastructure changes for potential speed limit increases. The objective is to drive quick wins by identifying bottlenecks and developing real-time and production ready tools. In one study, we analyzed train consist, speed, and topology data patterns to quantify train stall risk. The proposed methodology is used to enhance the existing tonnage rating methodology into a dynamic approach that can both prevent train stalls and unlock tonnage capacity by utilizing track topology features. Accurate speed limit is critical in optimizing network velocity and fuel efficiency. In another study, a track topology scanning algorithm is designed to search, simulate, and propose new speed limits with potential fuel and time savings and minimum capital investment. The process first filters out segments conform to certain geometry patterns (e.g., valley), where train speed can benefit from gravity. The candidate solutions are then analyzed in an automated simulation and statistical analysis pipeline to generate a final short list optimized to meet engineering team's criteria.

3 - Or Applications in Intermodal Operations**Gita Ayu, Norfolk Southern, Atlanta, GA, United States**

Intermodal transportation is a strategic approach that maximizes the strengths of each transport mode by loading cargo into containers, enabling seamless movement between modes. The life cycle of a shipment begins at ingate at the origin and concludes at outgate at the destination, utilizing terminal resources to store and transfer shipments within and between intermodal terminals. As one of the leading players in the intermodal industry, Norfolk Southern leverages its extensive network of terminals, access to major East Coast container ports, and partnerships with North American intermodal rail providers to ensure efficient and reliable goods movement. At Norfolk Southern, we develop, maintain, and continuously enhance our in-house, data-driven tools to support operations. These tools include shipment forecasting, railcar tracking, and terminal resource optimization. We will present an overview of each tool with their use cases.

4 - Operating Plan Optimization**Gunnar Feldmann, Norfolk Southern, Atlanta, GA, United States**

The operating plan drives all car movements at Norfolk Southern. It is revised on a daily basis by the Network Design Optimization group (NDO) to ensure the most efficient routing of cars throughout the network. NDO has several optimization tools at their disposal to achieve this goal. At the traffic side, Market, Research, and Forecast (MR&F) provides a judgemental as well as a model based volume forecast. This forecasted volume feeds into the offline plan developer tool to simulate the traffic movement based on the blocking and train plan. To achieve the most efficient routing and to sustain a continuous refinement of the operation plan, NDO utilizes additional decision support tools. For the blocking plan a very large scale neighborhood search heuristic model has been deployed, and for the train plan a heuristic routing model has also been ran. In this presentation, we will focus on the afore-mentioned models and, more specifically, how they contribute to an optimized operating plan.

MC21

Building A Level 4 A407

TSL Best Paper Award and TS Journal Paper of the Year award

Award Session

Transportation Science and Logistics (TSL)

Chair: Michal Tzur, Tel Aviv University, Tel Aviv, Israel

1 - Drone-Delivery Network for Opioid Overdose: Nonlinear Integer Queueing-Optimization Models and Methods**Miguel Lejeune, George Washington University, Washington, DC, United States, Wenbo Ma**

This presentation will be given by a winner of a TSL award.

2 - Routing Optimization with Vehicle - Customer Coordination**Alexandre Jacquillat, MIT Sloan School of Management, Cambridge, MA, United States, Wei Zhang, Kai Wang, Shuaian Wang**

This presentation will be given by a winner of a TSL award.

3 - D-Optimal Orienteering for Post-Earthquake Reconnaissance Planning**Jiaqi Wang, University of Maryland, College Park, MD, United States, Weijun Xie, Ilya Ryzhov, Nikola Markovic, Ge Ou**

This presentation will be given by a winner of a TSL award.

4 - Economic Analysis of On-Street Parking with Urban Delivery**Zhengtian Xu, George Washington University, Washington, DC, United States, Xiaotong Sun**

This presentation will be given by a winner of a TSL award.

MC22

Building A Level 4 A408

Decision-making in Structured Stochastic Systems

Invited Session

Applied Probability Society

Chair: Weina Wang, Carnegie Mellon University, Pittsburgh, PA, United States

1 - New Approaches for Restless Bandits and Weakly Coupled Mdp**Yige Hong, Carnegie Mellon University, Pittsburgh, PA, United States, Weina Wang, Qiaomin Xie, Yudong Chen, Xiangcheng Zhang**

In this talk, we present recent progresses on the classic restless bandit (RB) problem and its generalization, weakly-coupled Markov decision processes (WCMDP), under the long-run average-reward objective. These problems model the resource allocation and control of a potentially large number of weakly-coupled “arms”, each with a state that evolves dynamically. The RB and WCMDP problems have a long history and wide range of applications, but they are challenging to solve due to the high-dimensional joint state space.

In our recent line of work, we introduce novel policies and analysis tools that pushes the existing asymptotic optimality results beyond the classic setting. This includes removing a long-standing global attractor assumption, and allowing fully heterogeneous arms. We also establish exponentially small optimality gaps for non-degenerate RBs, under weaker-than-standard conditions. Our approach highlights a novel way of exploiting the weakly-coupled structure, by matching the long-term behavior of some subsystems with simpler processes. This is in contrast with the classic index policies, which are mainly greedy policies with less control of the system’s long-term behavior.

2 - Exploiting Exogenous Structure for Sample-Efficient Reinforcement Learning

Jia Wan, Massachusetts Institute of Technology, Cambridge, MA, United States, Sean Sinclair, Devavrat Shah, Martin Wainwright

We study Exo-MDPs, a structured class of Markov Decision Processes (MDPs) where the state space is partitioned into exogenous and endogenous components. Exogenous states evolve stochastically, independent of the agent’s actions, while endogenous states evolve deterministically based on both state components and actions. Exo-MDPs are useful for applications including inventory control, portfolio management, and ride-sharing. Our first result is structural, establishing a representational equivalence between the classes of discrete MDPs, Exo-MDPs, and discrete linear mixture MDPs. Specifically, any discrete MDP can be represented as an Exo-MDP, and the transition and reward dynamics can be written as linear functions of the exogenous state distribution, showing that Exo-MDPs are instances of linear mixture MDPs. For unobserved exogenous states, we prove a regret upper bound of $\tilde{O}(H^{\frac{3}{2}}\sqrt{K})$ over KS trajectories of horizon H , with d as the size of the exogenous state space, and establish nearly-matching lower bounds. Our findings demonstrate how Exo-MDPs decouple sample complexity from action and endogenous state sizes, and we validate our theoretical insights with experiments on inventory control.

3 - Sample-Path Comparison of Fcfm and ROS in One- and Two-Sided Matching Queues

Runhan Xie, University of California, Berkeley, Berkeley, CA, United States, Elene Anton, Kristen Gardner, Rhonda Righter

We consider a matching model with bipartite compatibilities and Poisson arrivals of items to both sides. The arriving items may queue on one or both sides. We find conditions under which FCFM has better performance than random matching in terms of minimizing the number of waiting items in the transient regime. We also show that under the same conditions, the LQ (match the longest compatible queue) is optimal in a strong sense among policies that can use all of the state information.

4 - Optimal Single-Policy Sample Complexity and Transient Coverage for Average-Reward Offline RL

Matthew Zurek, University of Wisconsin-Madison, Madison, WI, United States, Guy Zamir, Yudong Chen

We study offline reinforcement learning in average-reward MDPs, which presents increased challenges from the perspectives of distribution shift and non-uniform coverage, and has been relatively underexamined from a theoretical perspective. While previous work obtains performance guarantees under single-policy data coverage assumptions, such guarantees utilize additional complexity measures which are uniform over all policies, such as the uniform mixing time. We develop sharp guarantees depending only on the target policy, specifically the bias span and a novel policy hitting radius, yielding the first fully single-policy sample complexity bound for average-reward offline RL. We are also the first to handle general weakly communicating MDPs, contrasting restrictive structural assumptions made in prior work. To achieve this, we introduce an algorithm based on pessimistic discounted value iteration enhanced by a novel quantile clipping technique, which enables the use of a sharper empirical-span-based penalty function. Our algorithm also does not require any prior parameter knowledge for its implementation. Remarkably, we show via hard examples that learning under our conditions requires coverage assumptions beyond the stationary distribution of the target policy, distinguishing single-policy complexity measures from previously examined cases. We also develop lower bounds nearly matching our main result.

MC23

Building A Level 4 A409

MIF Paper Competition

Award Session

Minority Issues Forum

Chair: Mike Gordon, Virginia Tech, Blacksburg, VA, United States

Co-Chair: Jessye Talley, Morgan State University, Baltimore, MD, United States

1 - Incentivizing Healthy Food Choices Using Add-on Bundling: a Field Experiment

Nymisha Bandi, McGill University, Montreal, QC, Canada, Maxime Cohen, Saibal Ray

How can retailers incentivize customers to make healthier food choices? Price, convenience, and taste are known to be among the main drivers behind such choices. Unfortunately, healthier food options are often expensive and infrequently promoted. Recent efforts in deploying healthy nudges to incentivize customers toward healthier food choices have been observed. In this paper, we conducted a field experiment with a global convenience store chain to better understand how different add-on bundle promotions influence healthy food choices. We considered three types of add-on bundles: (i) an unhealthy bundle (when customers purchased a coffee, they could add a pastry for \$1), (ii) a healthy bundle (offering a healthy snack as an add-on), and (iii) choice bundle (offering either a pastry or a healthy snack). In addition to our field experiment, we conducted an online lab study to strengthen the validity of our results. We found that offering healthy snacks as part of an add-on bundle significantly increased healthy purchases (and decreased unhealthy purchases). Surprisingly, this finding continued to hold for the choice bundle, that is, even when unhealthy snacks were concurrently on promotion. Unfortunately, we did not observe a long-term stickiness effect, meaning that customers returned to their original (unhealthy) purchase patterns once the healthy or choice bundle was discontinued. Finally, we show that offering an add-on choice bundle is also beneficial for retailers, who can earn higher revenue and profit.

2 - Differential Game Theoretic Models for Designing Water Conservation Incentives

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

The growing concern over water scarcity and the management of freshwater is common worldwide. Voluntary incentives, such as payments to water users, offer a potential strategy to reduce conflicts and water consumption. This study formulates non-cooperative differential game-theoretic framework to design incentive schemes for water users, both groundwater and surface water, located in a river basin network. A non-governmental organization (NGO) acts as a key player, providing conservation incentives that account for both vertical and horizontal interactions among water users. The model designs incentives based on the characteristics of each water user and addresses scalability challenges as the network size grows. We propose solution methods for both convex and nonconvex settings, including Karush-Kuhn-Tucker conditions, Value Iteration, and Basis Function Approximation. Parametric analyses assess how model parameters influence solution methods and decision-making outcomes.

3 - Optimal Resource Allocation to Minimize Errors when Detecting Human Trafficking

Abhishek Ray, George Mason University, Fairfax, VA, United States, Viplove Arora, Kayse Maass, Mario Ventresca

Accurately detecting human trafficking is particularly challenging due to its covert nature, difficulty in distinguishing trafficking from non-trafficking exploitative conditions, and varying operational definitions. Typically, detecting human trafficking requires resource-intensive efforts from resource-constrained anti-trafficking stakeholders. Such measures may need personnel training or machine learning-based identification technologies that suffer from detection errors. Repeated usage of such measures risks biasing detection efforts and reducing detection effectiveness. Such problems raise the question: “How should imperfect detection resources be allocated to most effectively identify human trafficking?” As an answer, we construct a class of resource allocation models that considers various optimal allocation scenarios. These scenarios range from optimal location selection for monitoring to optimal allocation of a finite set of imperfect resources, given error rates. We illustrate the applicability

of these models across both human and technology-facilitated detection contexts at the India–Nepal border and in the global seafood industry. Insights from our models help inform operational strategies for allocating limited anti-human trafficking resources in a way that effectively preserves human rights and dignity.

MC24

Building A Level 4 A410

Reinforcement Learning: Theory and Algorithms

Invited Session

Applied Probability Society

Chair: Zaiwei Chen, Purdue University, West Lafayette, IN, United States

1 - Non-Asymptotic Guarantees for Average-Reward Q-Learning with Adaptive Stepsizes

Zaiwei Chen, Purdue University, West Lafayette, IN, United States

This work presents the first finite-time analysis of average-reward Q -learning with an asynchronous implementation. A key feature of the algorithm we study is the use of adaptive stepsizes that act as local clocks for each state-action pair. We show that the mean-square error of this Q -learning algorithm, measured in the span seminorm, converges at a rate of $\mathcal{O}(\frac{1}{k})$. To establish this result, we demonstrate that adaptive stepsizes are necessary: without them, the algorithm fails to converge to the correct target. Moreover, adaptive stepsizes can be viewed as a form of implicit importance sampling that counteracts the effect of asynchronous updates. Technically, the use of adaptive stepsizes causes each Q -learning update to depend on the full sample history, introducing strong correlations and making the algorithm a non-Markovian stochastic approximation (SA) scheme. Our approach to overcoming this challenge involves (1) a time-inhomogeneous Markovian reformulation of non-Markovian SA, and (2) a combination of almost-sure time-varying bounds, conditioning arguments, and Markov chain concentration inequalities to break the strong correlations between the adaptive stepsizes and the iterates.

2 - Concentration of Variance-Reduced Stochastic Approximation Algorithms

Bulut Boru, University of Minnesota, Minneapolis, IN, United States, Sajad Khodadadian, Martin Zubeldia

In this talk, we consider several variance-reduction techniques for stochastic approximation algorithms, and we go beyond variance by exploring their impact on the distribution of the error. In particular, we show that adding a projection step onto a ball with a slowly increasing radius can lighten the tail of the error from heavy-tailed (up to Pareto) to Gaussian.

3 - What Mathematical Structures Make RL Possible? An Elliptic Theory for Controlled Diffusions

Wenlong Mou, University of Toronto, Toronto, ON, Canada

Reinforcement learning (RL) offers a powerful framework of decision-making in dynamical environments, and recent years have seen a surge of RL applied to continuous-time Markov diffusions. Yet, existing contraction-based RL theory suffers from limitations such as diverging effective horizons, and fails to provide the relevant guidance.

In this talk, I will present recent advances in designing RL algorithms for controlled diffusion processes. I begin by exploring how *uniform ellipticity*—a foundational property of Markov diffusions—enables RL with rigorous theoretical guarantees. Among other findings, I will demonstrate how elliptic structures makes RL as easy as supervised learning, a striking contrast to the inherent difficulties of RL in unstructured Markov decision processes.

4 - $O(1/K)$ Finite-TIME Bound for Non-Linear Two-Time-Scale Stochastic Approximation

Siddharth Chandak, Stanford University, Stanford, CA, United States

Two-time-scale stochastic approximation is an algorithm with coupled iterations which has found broad applications in reinforcement learning, optimization and game control. While several prior works have obtained a mean square error bound of $O(1/k)$ for linear two-time-scale iterations, the best known bound in the non-linear contractive setting has been $O(1/k^{2/3})$. In this work, we obtain an improved bound of $O(1/k)$ for non-linear two-time-scale stochastic approximation. Our result applies to algorithms such as gradient descent-ascent and two-time-scale Lagrangian optimization. The key step in our analysis involves rewriting the original iteration in terms of an averaged noise sequence which decays sufficiently fast. Additionally, we use an induction-based approach to show that the iterates are bounded in expectation.

MC25

Building A Level 4 A411

Truck Platooning and Vehicle Routing Problems

Invited Session

TSL: Freight Transportation

Chair: Rui Zhang, University of Colorado Boulder, Boulder, CO, United States

1 - Container Drayage Operations with Autonomous Trucks and Platooning

Ziyu Qiu, University of Illinois Chicago, Chicago, IL, United States, Bo Zou, Chi Xie, Mingzhou Jin

Drayage plays a critical role in the national supply chain. While accounting for a small portion of the total distance traveled, drayage contributes a disproportionately large share of the total logistics cost. The connected and autonomous vehicle technologies presents opportunities to enhance the cost efficiency of drayage, but has not been well studied. This research evaluates the impact of autonomous trucks equipped with platooning capabilities on container drayage between intermodal terminal facilities. We develop a series of mixed integer programming models to characterize truck and container movements in time-expanded networks, under multiple scenarios including single (truck type) fleet without platooning, single fleet with platooning, mixed fleet with homogeneous platooning (trucks in a platoon are of the same type), and mixed fleet with heterogeneous platooning. For the last scenario, a linear programming-based approach is further embedded for optimizing truck composition while forming platoons. Extensive numerical experiments are conducted in the context of container drayage in the Chicago metropolitan area. Among the many results, we find that forming smaller platoons are preferred over larger platoons when the energy saving percentages are comparable. In addition, heterogeneous platooning generates greater more benefits than homogeneous platooning. The findings help inform future drayage system operation strategies with autonomous trucks.

2 - Electric-Truck Routing and Platooning with Capacitated Charging Stations

Danesh Hosseinpahani, University of Illinois-Chicago, Chicago, IL, United States, Jialu Yang, Bo Zou, Jane Lin

The integration of truck platooning and electric trucks presents a promising avenue for improving operational efficiency and environmental sustainability in freight transportation. Realizing the energy and cost saving as well as emission reduction benefits requires a holistic design of truck routing, scheduling, and platooning strategies that accounts for practical operational constraints. This study investigates the integrated problem of electric-truck routing and charging scheduling, incorporating key factors such as energy savings from platooning, battery limitations of electric trucks, charging station availability and capacity, mandatory driver rest periods, and the positional configuration of trucks within a platoon (i.e., leaders and followers). The objective is to minimize the total charging cost, travel time cost, and waiting time costs of the system, thereby ensuring timely arrivals across multiple origin–destination (OD) pairs, while simultaneously maximizing the formation of energy-efficient platoons along the routes. A mixed-integer linear program (MILP) is formulated to optimize truck routing, charging, and platooning decisions, while satisfying operational constraints. To solve the problem, we first use Gurobi to solve instances on a small hypothetical network. Then, we propose a Lagrangian Relaxation-based decomposition approach to seek near-optimal solutions for large-scale problems on the Illinois highway network. Extensive numerical experiments are conducted to examine the effectiveness and efficiency of the model and solution approach. The findings offer valuable managerial insights into optimizing electric-truck operations with platooning

3 - Planning of Truck Platooning for Road-Network Capacitated Vehicle Routing Problem

Zhibin Chen, New York University Shanghai, Shanghai, China, People's Republic of, Yilang Hao, Xiaotong Sun, Lu Tong

In real-world logistics, each truck may need to serve multiple customers located at different places, and the operator managing a fleet of trucks thus has to determine not only the routing and time schedules of each truck but also the set of customers allocated to each truck and their sequence to visit. This is well known as a capacitated vehicle routing problem with time windows (CVRPTW), and considering the application of truck platooning in such a problem entails new modeling frameworks and tailored solution algorithms. In light of this, this study makes the first attempt to optimize the truck platooning plan for a road-network CVRPTW in a way to minimize the total operation cost, including vehicles' fixed dispatch cost and energy cost, while fulfilling all delivery demands within their time window constraints. Specifically, the operation plan will dictate the number of trucks to be dispatched, the set of customers, and the routing and time schedules for each truck. In addition, the modeling framework is constructed based on a road network instead of a traditional customer node graph to better resemble and facilitate the platooning operation. A 3-stage algorithm embedded with a "route-then-schedule" scheme, Dynamic Programming, and Modified Insertion heuristic, is developed to solve the proposed model in a timely manner. Numerical experiments are conducted to validate the proposed modeling framework, demonstrate the performance of the proposed solution algorithm, and quantify the benefit brought by the truck platooning technology.

4 - A Two-Dimensional Vehicle Loading Problem with Heterogeneous Fleet and Delay Minimization

Juan morande, Auburn University, auburn, AL, United States, Juan Carlos Pina-Pardo, Mario Velez, Daniel Silva, Alice Smith

The approach presented in this paper not only increases the efficiency of existing delivery networks but also quantifies the effects of considering two dimensions when evaluating fleets and delivery routes. By solving the proposed problem in a timely manner, decision makers can plan deliveries along fixed routes on a daily basis. By showing the value of a two-dimensional approach in this case, the paper also contributes to the literature by showing the value of exploiting all the dimensions available in packing problems. Also, the paper gives a benchmark on the proposed problem for future work and methods. Finally, the paper also brings to attention the need for a rehandle minimization strategy when designing delivery networks. While in one-dimensional problems the optimal order in which trucks should be loaded is trivial and easily decided by any operator, when using two dimensions this problem requires a systematic approach to be solved.

5 - Effectation: An Exact Solution Approach for the Formation and Routing of Worker Teams

Rui Zhang, University of Colorado Boulder, Boulder, CO, United States, Jiliu Li, Jinfeng Yang, Nan Huang

Real-world applications often require tracking resource usage over time. One such application is the Worker Team Formation and Routing Problem (WTFRP), proposed by Dall'Olio and Kolisch (2023), which aims to improve airport ground-handling operations. The WTFRP involves scheduling multiple trips with time windows, forming teams with hierarchical and downgrade-compatible skills, and accounting for team-dependent task durations. A major challenge lies in tracking worker availability by skill level across the entire planning horizon, which is difficult for traditional mathematical programming approaches. Rather than discretizing time, this paper proposes an exact solution method in continuous time. We model WTFRP as an infinite integer program with infinitely many variables and constraints. To address its LP relaxation, we propose Effective Time-Point Detection (EFFECTION), a method that dynamically identifies time points to generate both variables and constraints. Building on EFFECTION, we develop Enumeration-Check-and-Switch (ECS). ECS includes: Enumeration, which fixes promising variables; Check, which verifies solution feasibility; and Switch, which alternates between the original and relaxed models. EFFECTION is embedded within ECS to guide the detection of effective time points. In the computational experiments on a set of known instances, we show that ECS not only produces better solutions but also achieves them significantly faster, up to two orders of magnitude, compared to the existing method. Furthermore, we highlight the capability of EFFECTION in solving problems in continuous-time settings on a new set of instances. The results underscore the importance of considering finer time resolutions, which leads to higher-quality solutions for airport ground-handling operations.

MC26

Building A Level 4 A412

Importance Sampling: Methods and Applications

Invited Session

Simulation

Chair: Dohyun Ahn, CUHK, Shatin, Hong Kong

1 - New Bounds and Truncation Boundaries for Importance Sampling

Yijuan Liang, Southwest University, Chongqing, China, People's Republic of, Guangxin Jiang, Michael Fu

Importance sampling (IS) is a technique that enables statistical estimation of output performance at multiple input distributions from a single nominal input distribution, commonly used in Monte Carlo simulation for variance reduction and in machine learning applications for reusing historical data, but its effectiveness can be challenging to quantify. We establish a new result showing the tightness of polynomial concentration bounds for classical IS likelihood ratio (LR) estimators. To address potentially high variance, we propose new truncation boundaries when using a truncated LR estimator, establishing upper concentration bounds that imply an exponential convergence rate. Numerical experiments illustrate the contrasting convergence rates of the various LR estimators and the effectiveness of the newly proposed truncation-boundary LR estimators for examples from finance and machine learning.

2 - Reusing Historical Trajectories in Natural Policy Gradient via Importance Sampling

Yuhao Wang, Georgia Institute of Technology, Atlanta, GA, United States, Yifan Lin, Enlu Zhou

Reinforcement learning provides a mathematical framework for learning-based control, whose success largely depends on the amount of data it can utilize. The efficient utilization of historical trajectories obtained from previous policies is essential for expediting policy optimization. Empirical evidence has shown that policy gradient methods based on importance sampling work well. However, existing literature often neglect the interdependence between trajectories from different iterations, and the good empirical performance lacks a rigorous theoretical justification. In this paper, we study a variant of the natural policy gradient method with reusing historical trajectories via importance sampling. We show that the bias of the proposed estimator of the gradient is asymptotically negligible, the resultant algorithm is convergent, and reusing past trajectories helps improve the convergence rate.

3 - Wasserstein Distributionally Robust Rare-Event Simulation

Dohyun Ahn, The Chinese University of Hong Kong, Shatin, Hong Kong, Lewen Zheng

Classical rare-event simulation methods mostly rely on precise knowledge of the underlying probability distributions, which limits their effectiveness under distributional uncertainty. To address this limitation, we develop a novel simulation framework for estimating rare-event probabilities under such distributional uncertainty. Specifically, our approach computes a distributionally robust rare-event probability bound over a family of distributions defined by a Wasserstein ball around a nominal Gaussian model. By exploiting a dual characterization of this bound and the concept of conditional importance sampling, our framework introduces a two-stage estimation procedure that is simple to implement. We prove the strong efficiency of the proposed method and demonstrate its superior numerical performance compared to existing approaches.

4 - Importance Sampling for Latent Dirichlet Allocation

Ayeong Lee, Columbia University, New York, NY, United States, Paul Glasserman

Latent Dirichlet Allocation (LDA) is a popular method in Natural Language Processing (NLP) for finding topics in text data. Evaluating an LDA model entails estimating the expected likelihood of held-out documents. This is commonly done through Monte Carlo simulation, which is prone to high relative variance. We propose an importance sampling estimator for this problem and characterize the theoretical asymptotic statistical efficiency it achieves in large documents. We illustrate the method in simulated data and in a dataset of news articles.

5 - Start Safe: Configuring Optimization Algorithms for Decision-Making Under Extreme Risks

Wasin Meesena, Columbia University, New York, NY, United States, Henry Lam

We consider stochastic optimization where the goal is not only to optimize an average-case objective, but also mitigate the occurrence and impact of extreme catastrophic events. This problem is motivated from safety-aware decision-making and AI training. In particular, in the presence of a simulation model, variance reduction techniques are naturally employed to control estimation errors due to event rarity. We argue, however, that natural attempts to integrate variance reduction into optimization, even executed in a reasonable adaptive fashion, encounters fundamental challenges in terms of guaranteeing realistic runtime when using common stochastic gradient descent algorithms. On a high level, the challenge arises from the extreme sensitivity of tail-based objectives with respect to the decision variables, which renders the failure of traditional Lipschitz-based analyses. We offer remedies based on a notion of "safe initialization", combined with careful update iterations, that allow for finite-time error control. We discuss implications of our findings on safe decision search and extremal predictive modeling.

MC27

Building B Level 2 B201

Emerging Applications of Optimization

Invited Session

Computing Society

Chair: Taghi Khaniyev, Amazon | Bilkent University, Bellevue, WA, 98005, United States

1 - The Inverse Optimal Value Problem for Linear Fractional Programming

Sina Nadi, University of Pittsburgh, Pittsburgh, PA, United States, Taewoo Lee, Oleg Prokopyev

We study the inverse optimal value problem for linear fractional programming, where the goal is to find the coefficients of the fractional objective function such that the resulting optimal objective function value is as close as possible to some given target value. We show that this problem is NP-hard. Then, we provide some structural results, which are exploited to derive several reformulations and two solution algorithms. The proposed approaches are based on the Charnes-Cooper and parametric transformations.

2 - Optimizing over Graph Neural Networks with An Application to Optimal Brain Tumor Resection

Kaan Çakıroğlu, Bilkent University, Ankara, Turkey, Taghi Khaniyev, Ozlem Karsu, Sahin Hanalioglu

Low-grade gliomas (LGGs) present a critical challenge in neurosurgical oncology due to their slow progression and the brain's adaptive neuroplastic reorganization. While complete resection reduces recurrence risk, it may disrupt reorganized functional networks. Emerging evidence shows partial resection can preserve neurological outcomes without compromising tumor control. However, surgeons lack quantitative tools to preoperatively evaluate these trade-offs. The absence of a computational framework to model resection trade-offs leaves surgical planning reliant on subjective intraoperative assessments rather than predictive network analysis. Our research addresses this gap by formulating the brain tumor resection problem as a mathematical optimization model using graph neural networks (GNNs) where the brain is represented as a multiplex network comprising regions of interest (ROIs) with structural and spatial connectivity. In our formulation, GNNs approximate brain functionality metrics such as global efficiency and modularity, serving as surrogate objective functions within a mixed-integer programming framework. We extend this methodology to bi-objective optimization, systematically analyzing trade-offs between several brain functionality metrics. Our framework provides surgeons with data-driven strategies that balance maximal tumor control with minimal network disruption. It achieves 99% accuracy in predicting postoperative functionality and efficiently approximates the Pareto Frontier of alternatives, enabling precision neurosurgery tailored to individual brain networks.

3 - An Optimization Approach to White Matter Brain Tumor Resection

Taghi Khaniyev, Bilkent University, Ankara, Turkey, Sena Asli Bozkurt, Sahin Hanalioglu, Oya Ekin Karasan

Preserving brain functionality while performing maximal tumor resection remains a central challenge in clinical practice, as current approaches lack a systematic framework to balance tumor removal with functional outcomes. To address this, we propose a mathematical optimization model that preserves network efficiency by resecting a predefined tumor volume. Network efficiency is quantified using the Global Efficiency (GE) metric. To reflect surgical feasibility, the model incorporates spatial contiguity constraints, ensuring that the resected region forms a connected area. To improve scalability, we developed enhanced formulations and an aggregation heuristic that groups physically close nodes, substantially reducing solution times while maintaining solution quality. Our experiments used real brain network data representing healthy white and gray matter regions, with synthetic tumor instances to simulate disruption. First, using a real tumor instance, we compared the "best" resection guided by our model with the "worst" resection, and demonstrated that our approach leads to significantly better preservation of brain network efficiency. We then validated the aggregation heuristic, showing that it drastically reduces computational time while achieving GE values comparable to the original model. Finally we analyzed the impact of tumor characteristics. Larger tumors were found to disrupt the network more severely even before resection, and to yield lower GE values after resection at equal threshold levels. Centrally located tumors also caused greater reductions in GE compared to peripheral tumors of the same size.

4 - A Hybrid Quantum-Classical Approach for Semidefinite Programming Problems

Gizem Ozbaygin, Bilkent University, Ankara, Turkey, Burak Kocuk, Diego Moran

Semidefinite programming plays a key role in optimization, with broad impact across fields like control theory, machine learning, and combinatorial optimization, thanks to its ability to capture complex problems in a tractable, convex framework. Although semidefinite programming problems are polynomially solvable, classical (non-quantum) algorithms can run into difficulties when dealing with large and dense matrices. Majority of the existing algorithms rely on iteratively evaluating positive semidefiniteness of a series of matrices. This is typically the bottleneck of the solution procedure because the algorithmic routines used for checking positive semidefiniteness mostly have cubic scaling with respect to the number of rows (or columns) in the matrices under consideration. In this study, we propose a hybrid quantum-classical framework to solve semidefinite programming problems by deploying quantum computing for the task of evaluating positive semidefiniteness while performing the rest of the computation classically. We present our preliminary findings based on an initial computational study.

5 - A Metamodel-Based General-Purpose Calibration Tool for Simulation Models

Elif Sena Isik, Georgia Institute of Technology, Atlanta, GA, United States, Taghi Khaniyev, Turgay Ayer, Jagpreet Chhatwal, Ismail Fatih Yildirim

Background: Simulation calibration is the process of configuring a simulation model's parameters to improve the agreement between the model output and the observed historical data. For most realistic simulation models, this calibration process can be quite computationally expensive, as it requires running the simulation model for each parameter combination. The inefficient use of computational resources is also a commonly encountered problem in medical decision making. To alleviate this problem, metamodels offer a trade-off between accuracy and computational efficiency for extensive simulative analysis with highly complex parameter space. **Method:** In this study, we propose an efficient metamodel-based calibration tool that any simulation-based study with computational challenges can benefit from. Four approaches (simulation-based calibration, metamodel-based calibration, and two hybrid methods that utilize the simulation and metamodel jointly) were tested to obtain the most effective tool in terms of accuracy and computational cost. **Results:** Metamodel-based calibration result was found to be an accurate indicator of the regions that are likely to provide optimal parameter configuration with a highly significant (approximately 98%) decrease in computational cost. Furthermore, the hybrid approach that utilizes metamodel to limit the search space and simulation to find the optimal configuration was found to have approximately 40% more precise results compared to the simulation-based calibration approach. **Conclusions:** The study displays the potential benefits of metamodeling in the simulation calibration process and offers a metamodel-based tool for calibration that can aid any simulation study with high computational cost.

MC28

Building B Level 2 B202

Synergies Between AI and Optimization

Invited Session

Computing Society

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

1 - Hyperplane Trees as Piecewise Linear Surrogates in Engineering Optimization

Carolina Colombo Tedesco, Carnegie Mellon University, Pittsburgh, PA, United States, Ethan Sunshine, John Kitchin, Carl Laird

Advances in Mixed-Integer Programming (MIP) have led to a growing interest in surrogate modelling. In this paradigm, nonlinear relationships between variables are "learned" by fitting a highly flexible function to some data samples. This fitted function must be one that can be formulated in an optimization problem, such as piecewise linear. To date, numerous piecewise linear surrogate modeling architectures have been explored, including ReLU neural networks, decision trees, and linear model decision trees (LMDT)s.

In this presentation, we introduce Hyperplane Trees (HT) as a new regression architecture designed to be embedded in optimization problems. An HT is a decision tree with linear responses in the leaves and linear combinations of variables used for splitting. HTs are more expressive than LMDTs, allowing better accuracy while introducing fewer binary variables. As an extension of decision trees, HTs are very fast and convenient to train on large or small datasets compared to ReLU neural networks.

We will present our implementation of HTs and show examples of their use in engineering optimization problems.

2 - Rigorous Model Uncertainty in Machine Learning Surrogates for Optimization

Carl Laird, Carnegie Mellon University, Pittsburgh, PA, United States, Daniel Varela

Machine learning (ML) surrogates are often embedded in optimization models to replace costly simulations, enable inverse design, and integrate black-box computations into solver-compatible formulations. While effective, these surrogates inevitably introduce approximation errors that can lead to infeasible or unreliable solutions. Recent advances in statistical learning allow to mitigate this issue by quantifying model uncertainty, but most existing methods do not provide rigorous guarantees when ML-derived constraints are integrated into optimization problems.

We propose Conformal Mixed-Integer Constraint Learning (C-MICL), a framework that delivers probabilistic feasibility guarantees for optimization problems with data-driven constraints. Unlike standard mixed-integer constraint learning methods, which may produce systematically infeasible solutions due to model error or limited data, C-MICL leverages conformal prediction to ensure ground-truth feasibility with probability at least $1-\alpha$, under a mild conditional independence assumption. The method accommodates both regression and classification tasks, requires no direct access to the true constraint function, and avoids the scalability bottlenecks of ensemble-based heuristics. Empirical results across real-world applications confirm that C-MICL consistently achieves target feasibility levels, retains strong objective performance, and reduces computational cost relative to existing approaches.

This work helps bringing mathematical optimization and modern ML uncertainty quantification by introducing a principled approach to enforce feasibility guarantees in surrogate-based optimization. Beyond improving robustness, C-MICL highlights how conformal prediction can be systematically integrated into optimization frameworks, enabling more reliable, data-driven decision-making in engineering and operations research.

3 - Integer Programming Formulations of Constrained Inference for Probabilistic Graphical Models

William Hart, Sandia National Laboratories, Albuquerque, NM, United States, Connor Mattes, Frank Qiu

Although probabilistic graphical models (PGMs) have been successfully applied to a wide range of problems, the analysis of many applications naturally needs to account for constraints that are cannot be directly represented with a PGM. For example, in cybersecurity applications the knowledge state of an attacker influences its ability to execute steps in an attack and associated log events. We describe the use of integer programming (IP) formulations to support constrained inference for probabilistic graphical models. IP formulations that have been proposed to compute the maximum a posteriori states of random variables in PGMs. We show how these formulations can be extended to support constrained inference, using simple exemplars to illustrate the potential application of this idea.

MC29

Building B Level 2 B203

Open Energy Modeling

Invited Session

ENRE: Electricity

Chair: Harry van der Weijde, Open Energy Transition, Bayreuth, 2274GA, Germany

1 - Open Modeling in High-Stakes Planning Processes

Harry van der Weijde, Open Energy Transition, Bayreuth, Germany

The transition to a low-carbon energy system demands robust planning tools capable of addressing increasing variability, uncertainty, and the growing importance of system flexibility. In both regulatory and corporate environments, energy modeling is increasingly expected to be not only technically rigorous but also open, transparent, and reproducible. Yet applying open modeling principles in mission-critical contexts poses unique challenges.

This talk presents practical insights from developing open energy models designed for high-stakes applications, including regulatory decision-making in Europe and strategic planning processes at system operators. We focus on two core tensions encountered in this work. First, the challenge of reconciling openness with the need to manage highly sensitive data. This requires the implementation of strict data security protocols, such as isolated computing environments, tiered access controls, and anonymization procedures, without compromising the model's transparency or its accessibility to diverse stakeholders. Second, we address the technical and organizational demands of scalability. Mission-critical modeling often involves running thousands of complex scenarios under tight deadlines, which necessitate cloud computing implementations, but without sacrificing ease of use.

2 - PyPSA on the Cloud: Easy on-Demand Scaling for Open Energy Models

Soham Garg, AWS, London, United Kingdom, Satheesh Maheswaran

Many real-world energy planning studies require computational power that is beyond the reach of most laptops and desktops. Advanced studies also require running and analyzing large batches of scenarios in parallel. This is challenging for users who do not have access to institutional high performance computing clusters. This talk presents a cloud workflow for the PyPSA energy modelling framework, in the form of ready-to-use templates and infrastructure-as-code, that allows new and advanced PyPSA modellers to run their scenarios in the cloud, using per-hour and on-demand cloud computing. By simplifying access to cloud computing resources, this workflow enables more modellers to push the limits on larger, faster, and more accurate large-scale energy models.

3 - Energy Modeling with Open Data

Austen Sharpe, Catalyst Cooperative, Philadelphia, PA, United States, Christina Gosnell

Catalyst Cooperative created and maintains the Public Utility Data Liberation (PUDL) project to support a variety of decarbonization projects with free and open-source data about the U.S. electricity system. PUDL is used by a variety of users from academics to clean energy developers and advocates. One key use case of PUDL is energy modeling. PUDL is a core data input to several open source energy models including GenX, PyPSA-US and GridPath.

There is a wealth of available data about the energy system in the U.S. but it can be challenging to access and use. Data is published across dozens of websites in disparate challenging formats, there is often a lack of shared keys linking assets and entities and there are consistent gaps or inconsistencies in reporting.

Curating data used for many different models - as well as all of the other use cases for PUDL - presents unique challenges. There are many trade-offs we have to manage when transforming the original data to make it more useful for users. We aim for data that is updated regularly, easy to use, includes long time series, has complete geographic coverage, is clean and free of errors, is as granular as possible and is connected to other datasets. Most raw US datasets are not published with those attributes. While many energy modelers have many shared input data needs, their priorities vary greatly. Catalyst tries to manage these needs by compiling high-value and non-controversial data transformations to benefit all users.

4 - The Impacts of Spatial Scale in Disjoint Generation and Transmission Planning

Kamran Tehrani, Stanford University, Stanford, CA, United States, Inês Azevedo

In restructured electricity markets, transmission planning typically follows a “reactive” approach: generation investment plans are an exogenous input to transmission planning models. This sequential process limits planners from co-optimizing complementary generation and transmission investments. Furthermore, generation planners typically use simplified representations of transmission resolution and power flow as compared to transmission planners. We quantify the impacts of this mismatch in transmission representation within reactive and proactive transmission planning processes using a model of the ERCOT interconnection. Our findings reveal that the economic value of transmission information ranges from 4 to 11% of system costs under business-as-usual scenarios, highlighting the importance of proper characterization of transmission constraints and proactive planning.

5 - Open Modeling with ML: Exploring a Physics-Informed Flow Matching Approach for Accelerated Energy Network Optimization

Kshitiz Khanal, Institute for Transportation Research and Education (NC State University), Raleigh, NC, United States

Machine Learning (ML) techniques have been increasingly adopted and explored in energy modeling, because of the capabilities to support decision-making by modeling complex data-driven systems, and provide a faster throughput compared to legacy modeling techniques. The openness in ML community lends itself to the increasing adoption. This exploratory research contributes to open energy modeling by investigating "Flow Matching" (FM), a novel generative artificial intelligence technique, and test the promise of ML to accelerate power system optimization based on patterns in existing data.

This work focuses on developing an FM model that learns the underlying patterns of efficient power flow from a diverse range of simulated scenarios on standard test systems. The physics-informed loss functions ground the patterns learned from data to reality. The initial results are encouraging, showcasing the FM model's ability to produce near-optimal solutions faster than conventional DC Optimal Power Flow (OPF) solvers. Importantly, the model demonstrates a potential to generalize its predictions in previously unseen extreme conditions, such as significantly increased electricity demand or unexpected generator outages, often maintaining high solution quality and adherence to system constraints.

These findings underscore the potential of Flow Matching and machine learning in general as a useful, data-driven component within the open energy modeling toolkit. This approach could empower researchers and planners with faster, more agile decision-support capabilities, fostering wider participation and innovation in the design and operation of resilient and efficient future energy systems enabled by the openness in machine learning community of practice.

6 - Mosaic: Open-Source Agent-Based Infrastructure Planning

Ruaridh Macdonald, Massachusetts Institute of Technology, Cambridge, MA, United States, yu weng

Recent studies of decarbonized electricity systems have shown the importance of low-emission firm generation in reducing the average cost of electricity, particularly when considering extreme events and interannual variation of variable renewable generation demand. These studies of the value of firm generation have used optimization-based capacity expansion models to determine the optimal design and operation of the electricity grid. While several of these studies have used multi-period optimizations, they did not consider how the ordering of investments or short-term private incentives may lead to higher electricity prices and / or worse grid reliability. This was because optimization-based approaches to this analysis struggle to explore path-dependency effects, heterogenous and imperfect future information, and private interests which are not aligned with public ones.

As an alternative approach, we are developing Mosaic, an agent-based capacity expansion model, to simulate the progression of the energy transition and evaluate the role of different forms of clean firm generation in ensuring a resilient and reliable electricity grid. We believe Mosaic offers new insights into grid investment decisions, as the cost of less insights into energy markets and dispatch decisions than other agent-based grid models. This talk discusses our recent progress and the impact of developing Mosaic as an open-source project.

MC30

Building B Level 2 B204

Optimization and Market Design of Energy Storage in Power Systems

Invited Session

ENRE: Electricity

Chair: Dongwei Zhao, Argonne National Laboratory, Westmont, IL, United States

1 - Price Formation and Long-Term Equilibrium in Future Electricity Markets: the Role of Energy Storage

Audun Botterud, Massachusetts Institute of Technology, Cambridge, MA, United States, Magnus Korpås, Guillaume Tarel

As existing generation resources retire, Variable Renewable Energy (VRE) and Energy Storage (ES) will have an increasingly important impact on the price dynamics in future electricity markets. A key question is whether the expected future dominance of VRE and ES requires a complete rethink of electricity markets or if marginal improvements to existing market designs are sufficient. We use a stylized analytical model based on load duration curves to illustrate price formation, revenue sufficiency, and long-term market equilibrium under the transition towards zero-carbon electricity systems, with a particular focus on the role of ES. We find that energy-only markets with scarcity prices set to the value of lost load may continue to provide adequate incentives for operations and investments after the introduction of VRE and ES. Moreover, ES plays an increasingly important role in price dynamics as thermal plants are pushed out of the competitive equilibrium. If the system reaches a state where flexible generators with non-zero marginal costs are no longer part of the optimal resource mix, our results indicate that equilibrium prices need to directly reflect the capital costs of VRE and ES. Our results raise questions around which cost elements should be allowed in short-term energy market offers, the role of capacity remuneration mechanisms, and the importance of long-term energy contracts in zero-carbon markets dominated by VRE and ES.

2 - Reliability Assessment of Energy Storage as a Transmission Asset

Dongwei Zhao, Argonne National Laboratory, Lemont, IL, United States, Zhi Zhou, David Schloff

Energy storage technologies have been technically proven to be capable of providing transmission services. However, their unique operational constraints as transmission assets and technical complexity present a range of challenges in technical feasibility, economic viability, and regulatory alignment within existing ISO/RTO practices. A research gap exists in the literature and available tools for assessing the reliability contribution of energy storage as a transmission asset, which lays the foundation of cost-benefit assessment of energy storage-based transmission solutions.

This work develops an integrated simulation-based reliability-assessment framework to evaluate the reliability contribution of energy storage as a transmission asset, which is called to provide transmission service when a N-1-1 contingency occurs. The assessment is designed for two deployment scenarios: transmission-asset only and dual use where a storage project may participate in a wholesale market. We define the period when an energy storage project serves as a transmission asset as Transmission Service Period (TSP). In the transmission-asset only scenario, the TSP is the whole 24 hours period. In the dual-use scenario, the TSP period is estimated based on system reliability condition. Outside the TSP, energy storage is allowed market participation, with state-of-charge (SOC) constraints enforced to ensure availability during the TSP. We conduct sensitivity analyses on SOC levels and storage durations to assess their impact on both reliability performance and market revenues. By comparing the two scenarios, we quantify the economic and reliability values of dual-use strategies and offer insights to enhance the value proposition of storage as a transmission asset.

3 - Characterizing Real-Time Electric Vehicle Impacts: Integrated Transportation and Grid Modeling Under Extreme Weather

Siyuan Wang, Argonne National Laboratory, Lemont, IL, United States, YANTAO HUANG

As electric vehicle (EV) adoption increases, the storage capacity provided by these vehicles can enhance the resilience of grid operations. This work focuses on integrating transportation and power distribution grid models to improve urban resilience under extreme weather conditions. We incorporate detailed transportation simulations to account for weather-related delays, road closures, and congestion, while expanding EV charging behavior models to align with grid capabilities. The distribution grid model will be improved to support real-time grid recovery and leverage EVs for grid supports. The dynamic features from model integrations help minimize disruptions and optimize EV resource use for a more resilient infrastructure.

4 - Coupler-Compliant Pcm Chipps: Enabling Dynamic Co-Simulation of Hydropower and Grid Operations

Matija Pavicevic, Argonne National Laboratory, Lemont, IL, United States, Phumthep Bunnak, Stefano Galelli, Sean Turner, Quentin Ploussard

As hydropower's role in grid flexibility grows, the limitations of existing Production Cost Models (PCMs) become increasingly apparent. Traditional PCMs often oversimplify hydropower operations, failing to capture the physical and temporal constraints of water management. To address this, we are developing a high-performance, coupling-compliant PCM module—referred to as a PCM CHiPP (Coupler-Compliant High-Performance Plier)—as part of the HydroCHiPPs project.

The PCM CHiPP is a lightweight, open-source emulator of industry-standard PCM behavior, designed specifically for dynamic coupling with river operations models (ROMs). Built from the PowNet framework, the PCM CHiPP replicates short-term economic dispatch of hydropower units under price, demand, and system constraints. A primary design goal is to expose a well-defined API for two-way communication, enabling the exchange of dispatch targets, reservoir states, and system conditions at user-defined intervals (e.g., hourly, sub-daily).

This modular approach allows the PCM CHiPP to be integrated into coupled simulation environments without the overhead of full production-grade PCM software. The resulting system supports experimentation with synchronization strategies, data exchange frequency, and optimization timing—providing new insights into how ROM-PCM interactions affect dispatch efficiency, revenue potential, and system flexibility.

The PCM CHiPP is being configured for both a synthetic hydropower cascade test system, enabling reproducible benchmarking and publication. Through this work, we aim to provide a scalable tool that lowers the barrier to coupling adoption and helps bridge the gap between power system planning and hydropower reality.

5 - Differentiable Robust Optimization for Electricity Dispatch: a Gradient-Based Framework for Managing Uncertainty

Owen Shen, MIT, Cambridge, MA, United States, Hungpo Chao, Patrick Jaillet, Haihao Lu

Rising uncertainty in power systems—driven by renewable energy integration and extreme weather—challenges the effectiveness of traditional optimal dispatch methods. This paper presents a differentiable robust optimization framework for electricity dispatch with energy storage, leveraging duality theory, optimization theory, and gradient-based methods. The approach offers three key advantages: First, it accommodates time-varying uncertainty sets, allowing a more accurate representation of renewable generation volatility than fixed-set methods. Second, it supports fast re-optimization via warm-starting, enabling responsiveness to real-time forecast updates. Third, it leverages linear programming duality for efficient gradient computation, yielding economically meaningful shadow prices across uncertainty scenarios. A numerical case study illustrates the framework's scalability and adaptability, highlighting its potential to improve grid reliability and economic performance amid increasing uncertainty.

MC31

Building B Level 2 B205

Operations Research and Game-Theoretic Approaches for Sustainable Natural Resource and Energy Systems

Invited Session

ENRE: Natural Resources

Chair: Hyunwoo Lee, Virginia Tech, Blacksburg, VA, 24060, United States

1 - Game-Theoretic Modeling of Insurance Market Dynamics in Wildfire-Prone Regions

Jia Choi, University at Buffalo, Buffalo, NY, United States, Jun Zhuang

California faces increasing wildfire risks, especially in densely populated areas like Los Angeles County. Government regulations, such as premium caps, aim to maintain insurance affordability, but have led insurers to exit high-risk markets, leaving many homeowners uninsured. This study develops a game-theoretic model to examine strategic interactions among regulators, insurers, and homeowners. The model captures decisions related to market participation, regulation intensity, and insurance choices under varying wildfire risks. Using data from LA County, we calibrate payoffs based on insurance profits, regulation costs, and wildfire losses. Results show that overly strict regulation can destabilize the insurance market, while balanced policies with mitigation incentives support both affordability and market stability. The findings offer policy insights to prevent insurance inaccessibility in wildfire-prone areas under climate change.

2 - Optimizing Surveillance and Management of Round Goby Invasion in Lake Champlain

Sabah Bushaj, State University of New York at Plattsburgh, Plattsburgh, NY, United States

The invasive round goby (*Neogobius melanostomus*) is an imminent ecological and economic threat to Lake Champlain, approaching via the Champlain Canal and Richelieu River. With the significant economic impacts of biological invasions escalating, proactive and strategic management is urgently needed. In this study, we propose an integrated simulation-optimization framework explicitly designed to optimize surveillance, prevention, and control efforts within realistic budgetary and logistical constraints. The simulation component of our model provides a spatially explicit representation of round goby dispersal dynamics, integrating critical biological traits such as dispersal rates, reproduction patterns, and survival probabilities, alongside detailed hydrodynamic modeling of the Champlain Canal and Richelieu River. We simulate multiple scenarios representing different levels of infestation risk (low, medium, high) to comprehensively explore potential invasion outcomes. The simulation outputs feed into an optimization model, formulated as a mixed-integer programming problem. The optimization seeks to minimize ecological damage, management costs, and overall resource usage. The decision variables include targeted surveillance locations and frequencies, round goby removal operations, and operational decisions concerning canal lock management. Constraints embedded within the model ensure adherence to practical limitations such as available budgets, labor, and equipment resources. The results from computational experiments provide clear and actionable strategies for resource allocation and management operations, including optimized recommendations for surveillance and removal efforts, and specific operational guidelines for canal lock management. This integrated methodological framework offers a flexible, generalizable approach applicable to managing other invasive species and similar ecological threats in aquatic ecosystems.

3 - Spatial Optimization Models for Width-Constrained Wildlife Corridor Design

Amirreza Saffarian, University of Florida, GAINESVILLE, FL, United States, Jorge Sefair

Human activities are increasingly fragmenting natural habitats, placing many species at risk. This creates an urgent need to preserve biodiversity and maintain ecological connectivity through wildlife corridors. We present two spatial optimization models for corridor design that explicitly incorporate corridor width as a key ecological criterion. The first model minimizes total corridor cost while meeting minimum width requirements, and the second maximizes corridor width under a fixed budget. Together, these models offer a flexible framework for connecting multiple fragmented areas, accommodating both regular and irregularly shaped candidate patches. To improve computational efficiency for the width maximization model, we develop a binary search-based algorithm and introduce a cut separation technique that integrates spatial information with mathematical formulations to generate strong initial bounds on achievable corridor width. We evaluate the models and solution methods on a range of synthetic and real-world landscapes, including a conservation case study focused on maintaining habitat connectivity for the Florida panther.

4 - A Wildfire Propagation Graph-Based Approach for Firebreak Planning in Fire-Prone Landscapes

Denys Yemshanov, Canadian Forest Service, Ontario, ON, Canada, Ning Liu, Eric Neilson, Daniel Thompson, Frank Koch

Firebreaks and fuel treatments are critical means for the reduction of wildfire hazard in fire-prone landscapes. However, the uncertain behaviour of wildfires makes the planning of firebreaks challenging, especially when the available resources are insufficient to treat all locations under wildfire threat. We propose a fire propagation graph-based approach that uses acyclic directed graphs to track the possible spread of wildfires in complex landscapes and estimate the impacts of firebreak placement decisions on the potential burn area at the landscape-level scale. We integrate the fire propagation graph concept into a mixed-integer programming model that allocates firebreaks in a fire-prone landscape. We used this model to compare two firebreak planning strategies. The first strategy reduces the connectivity between forest patches with fuel and minimizes the number of location pairs between which a wildfire could spread in the landscape. The second strategy minimizes the potential burn area in the landscape by tracking the impact of the firebreaks on the fire spread through a set of fire propagation graphs. We compared optimization problems for each strategy with a problem that combines both strategies.

We illustrated the approach with the planning of wildfire mitigation measures in the Red Rock-Prairie Creek area of Canada, a fire-prone forest landscape. The firebreak solutions were able to effectively reduce both the potential burn area and the connectivity between sites with fuel in the landscape. Our solutions help uncover the trade-offs between the different firebreak planning strategies and could assist wildfire mitigation planning in other regions.

MC32

Building B Level 2 B206

Modeling the Energy Transition in Africa

Invited Session

ENRE: Energy-Climate

Chair: Erin Baker, Univ of Massachusetts-Amherst, Amherst, MA, United States

1 - An Equitable Energy Transition: Integrating Community Perspectives into Power System Modeling in an Emerging Economy

Faustina Sarfo, University of Massachusetts, Amherst, MA, United States

Ghana faces significant challenges in providing reliable and equitable electricity to its population. This energy crisis, especially in Northern Ghana, has resulted in unreliable power supplies that disproportionately affect marginalized communities, hindering economic growth and quality of life. Our ultimate goal is to integrate community-driven metrics into an energy system model for Ghana, ensuring that energy planning not only addresses technical challenges but also incorporates the needs of marginalized communities. As a first step to parameterizing the model, we will analyze electricity demand data from Northern Ghana to identify blackout periods and estimate latent demand using a statistical model. This will provide insight on how blackouts impact grid performance and offer a clearer understanding of the unmet electricity demand in underserved regions. As a next step, we will develop a version of GenX tailored specifically to Ghana's energy system, focusing on implementing the first metric of reliability-adjusted access. This metric, which combines electricity access with system reliability, will be integrated into GenX to assess energy investment pathways and their potential to improve both grid reliability and equitable access to electricity in rural regions. This will directly contribute to simulating energy system configurations and evaluating investment strategies that prioritize these equity goals. By linking the estimation of latent demand with the implementation of reliability-adjusted access in GenX, the research will help set the stage for future applications that prioritize equity in Ghana's energy transition.

2 - Modelling Political Economy Constraints in Nigeria's Power Sector: Tariff Structures, Subsidies and Investment Decisions

Oladoyin Phillips, Princeton University, Princeton, NJ, United States, Emil Dimanchev, Jesse Jenkins

Regulating the electricity sector in countries in the Global South requires acknowledging the need for a careful balance between political and economic incentives and the market constraints this balance produces. In Nigeria, the resulting policy environment features electricity tariffs that remain below revenue-sufficient levels, contributing to underinvestment in grid capacity and unreliable service. To maintain system functionality, the government provides subsidies that artificially reduce the effective cost of grid electricity for distribution companies (DisCos), enabling them to purchase power at rates below economically sustainable levels. Furthermore, the Service-Based Tariff (SBT) regime was introduced to allow higher tariffs in exchange for improved service quality. However, the combined effects of these measures on cost, reliability, and the energy mix are unclear. This study presents an electricity system equilibrium model that analyses how Nigeria's electricity subsidy regime and SBT structure influence investment and operational decisions by electricity producers and consumers. The model examines the DisCo's investment decisions and allocation of limited grid electricity across feeders and consumer decisions regarding distributed energy technologies—solar PV, batteries, and fossil fuel generators. Sensitivity tests on diesel, solar, and battery costs explore trade-offs in system reliability, affordability, and emissions intensity. These tests also reveal scenarios where grid defection is more likely or where alternative business arrangements involving a mix of on- and off-grid supply may be more appropriate. This approach offers an analytical framework for understanding second-best outcomes under political and financial constraints, supporting context-specific transition pathways for African electricity systems.

3 - Multi-Stage Optimization Models for Cost-Effective Rural Electrification Planning

Ayşe Selin Kocaman, Bilkent University, Ankara, Turkey, Gözde Yazici, Oya Ekin Karasan

Approximately 9% of the global population still lacks access to electricity, with most residing in rural areas. Addressing this challenge requires the careful planning of decentralized energy systems that balance cost, reliability, and scalability. In this study, we develop novel mathematical optimization models to support rural electrification by determining the optimal technology—grid, mini-grid, or off-grid—for each demand point, while simultaneously designing cost-effective grid and mini-grid networks.

We extend these models into a multi-stage framework that captures real-world constraints such as limited budgets and infrastructure rollout timelines, enabling phased implementation. This dynamic approach demonstrates clear advantages in cost and flexibility over static, single-stage planning methods.

Given the scale and complexity of the problem—an extension of the NP-hard k-MST problem—we incorporate advanced solution techniques, including decomposition methods and non-compact, cut-based formulations, to improve computational performance. Numerical experiments on both real-world and synthetic datasets demonstrate the effectiveness and scalability of our models across diverse rural settings.

Our findings provide a rigorous and practical decision-support framework for rural electrification planning. The proposed tools support broader development goals, particularly Sustainable Development Goal 7 and the World Bank's Mission 300 initiative, by enabling cost-efficient and scalable energy access strategies.

4 - Bridging Climate Risk and Grid Resilience in Africa's Energy Transition

Kwame Donkor, University of Washington, Seattle, WA, United States, June Lukuyu

As climate risks intensify and urban populations grow, strengthening the resilience of electricity distribution networks in Sub-Saharan Africa is increasingly urgent. This work presents modeling and statistical analysis approaches to support data-driven, equitable grid planning in resource-constrained contexts, with a focus on Accra, Ghana. We leverage a first-of-its-kind, high-resolution dataset on distribution-level outages and voltage quality in the region to investigate how extreme weather events, including heatwaves and intense rainfall, interact with grid reliability and power quality in Accra. Our results show that climate-vulnerable neighborhoods disproportionately experience long-duration outages and sustained undervoltages, compounding existing socioeconomic risks. By combining statistical analysis, machine learning, and stakeholder engagement, we demonstrate how low-cost sensing and open-source tools can help bridge the data-to-impact gap. This work contributes a replicable framework for identifying priority areas for resilience investment and for modeling how climate and social vulnerabilities intersect with infrastructure weakness. The approaches and insights presented aim to support more adaptive, inclusive energy system planning as the region moves toward electrification, decarbonization, and development goals.

5 - Participatory Policy Analysis of Electric Cooking for Informal Settlements in Kampala, Uganda

Civian Kiki Massa, University of Massachusetts, Amherst, MA, United States, Jessica Kersey

Air pollution from cooking with polluting biomass and fossil fuels causes an estimated 740,000 premature deaths annually in sub-Saharan Africa (SSA). Despite decades of effort, the uptake of transitional cooking fuels, such as liquefied petroleum gas, has been slow in the region. Electric cooking (e-cooking) is emerging as a viable alternative in urban areas due to the availability of grid infrastructure. However, applied research on efficient cooking appliances has largely overlooked the potential of e-cooking for vulnerable urban populations. Informal settlements, home to over 60% of SSA's urban population, disproportionately bear the burdens of polluting cooking fuels.

This study leverages policy benefit analysis modeling to quantify the missed opportunities of not prioritizing informal settlements in efficient e-cooking initiatives. Using data collected through participatory methods, we simulate the impacts of four scenarios of e-cooking uptake. Our analysis quantifies environmental benefits such as avoided emissions, health benefits including prevented premature deaths and averted disability-adjusted life years (DALYs), economic gains from fuel switching, and social benefits like time savings. By comparing these benefits to the costs of implementation, we provide actionable insights for policymakers seeking to balance equity and efficiency in energy transitions. Our findings emphasize the importance of integrating informal settlements into e-cooking strategies, highlighting their potential to advance sustainable development goals while improving the quality of life for underserved communities.

6 - Designing Decentralized Solar Systems in Dust Affected Regions of West Africa

Stewart Isaacs, Massachusetts Institute of Technology, Cambridge, MA, United States, Danielle Wood, Olivier de Weck, Wesley Harris

Decentralized solar energy is a promising strategy to expand modern energy access in rural West Africa. However, in many parts of the region, photovoltaic (PV) power output is reduced by seasonal dust aerosols, which both accumulate on panels (soiling) and attenuate incoming solar radiation. In this work, we develop a data-driven surrogate model to support the design of decentralized PV systems in dusty environments. The model takes as input environmental data such as particulate matter concentration and estimates PV energy generation, soiling losses, and system reliability. We use the model to explore how optimal system parameters—such as panel tilt angle and energy storage capacity—shift when accounting for environmental constraints, reliability, and revenue from energy services. Our findings highlight how dust effects can influence PV system design tradeoffs, and how modeling these effects can guide reliable deployments in under-resourced regions.

MC33

Building B Level 2 B207

Operational and Policy Innovations for Modern Healthcare

Invited Session

Health Applications Society

Chair: Liang (Leon) Xu, Singapore Management University, Singapore, Singapore

1 - Developing Health-Aware Built Environments for Personalized Healthcare

Yiwen Dong, Stanford University, Stanford, CA, United States

Our surrounding built environment, traditionally static and inert, is becoming an active partner in healthcare through advances in contactless sensing and artificial intelligence. This work introduces a new research concept of Health-Aware Built Environments (HABE) - spaces equipped with passive monitoring systems (i.e., unobtrusive, non-contact sensors such as vibration, radar, vision sensors) that continuously analyze human-environment interactions to uncover latent disease markers that guides the detection and intervention of personalized health conditions in daily living spaces. HABE captures traditionally hard-to-measure health signals from everyday behaviors, including gradual changes in mobility and sudden shifts in daily activity routines, that often go unnoticed in traditional clinical assessments. HABE enables proactive, precision, and personalized monitoring of various health conditions, such as neuropsychiatric, aging (e.g., Alzheimer's disease, Parkinson's, frailty), and neurodevelopmental (e.g., Autism) conditions. We present a framework for translating these daily behavior (e.g., room transitions, human-environment interactions) into disease markers, demonstrating how buildings can evolve from passive infrastructure to diagnostic agents for personalized healthcare. By bridging artificial intelligence, civil engineering, and clinical medicine, we envision a future where buildings themselves become intelligent allies in preventing and managing diseases, supporting aging in place, and fundamentally transforming our physical spaces into partners in personalized healthcare.

2 - Designing Specialist-Response Policies in Hospital Emergency Departments

Emily Zhu, Texas State University, San Marcos, TX, United States

Emergency department (ED) crowding continues to threaten timely care delivery worldwide. A major contributor to ED delays is the specialist consultation (SC) process, where ED physicians request expert evaluations from off-site specialists. These requests often face significant Request-to-Realization (R2R) delays—time gaps between request and consultation—due to specialists' competing responsibilities. Using data from over one million ED visits across 11 Montreal hospitals, we identify SC delays as a key bottleneck in patient flow, exceeding the impact of lab and imaging delays.

3 - Optimizing Family-Centered Genetic Testing Pricing and Product Portfolio

Ying Cao, Penn State University - Erie, Erie, PA, United States, Xinyu Wei, Yuchen Wang, Xianghui Peng, Victor Prybutok

The importance of a family-centered approach to genetic testing is gaining increasing recognition among healthcare practitioners. Genetic testing results within family units are crucial for selecting newer targeted therapies and prompting relatives to assess their own cancer risks, particularly for genetic disorders such as cancers. Despite these benefits, genetic testing remains underutilized by patients and their families, largely due to cost concerns, a shortage of genetic counselors, and limited or outdated genetics training among some healthcare providers. This study investigates the portfolio design and pricing of family-centered genetic testing product offerings in face of varied family structures and risk profiles. Our research provides recommendations for health service providers to design genetic testing product portfolio that maximizes health benefits for families and addresses the technological scalability.

4 - The "Netflix Model": A New Payment Model for Asymptomatic Disease Management

Huaiyang Zhong, Virginia Tech, Blacksburg, VA, United States

Rising drug prices challenge public payers aiming to expand access to curative therapies without destabilizing budgets. Two types of Subscription-Based Drug Contracts (SBDCs) have emerged to address this. In 2015, Australia adopted a pure subscription model, paying a fixed amount for unlimited hepatitis C treatment. In 2019, Louisiana and Washington launched "modified subscription" deals, paying per prescription up to a cap, after which additional treatment incurred no cost. Despite similar contracts, outcomes varied: Louisiana and Australia saw treatment surges, while Washington did not. This paper uses a game-theoretical model to explain this discrepancy, showing that SBDC effectiveness depends on both contract structure and the population's willingness to engage with screening. SBDCs are most impactful where

traditional outreach fails, as they incentivize identifying and treating hidden demand. We formalize this using a willingness-to-screen (WTS) metric, derive effectiveness thresholds, and calibrate the model with Medicaid data to provide targeted policy guidance.

5 - Reducing Prescription Errors Through Information Interventions: A Field Experiment in Healthcare Operations

Xiaodan Shao, Nanyang Technological University, Singapore, Singapore, Vivek kumar Choudhary, Arnab Majumdar, Burhanuddin Pithawala

Can simple, non-mandatory, and cost-effective informational interventions effectively reduce prescription errors, given that mandatory systems often fail due to high overrides by doctors? Do such interventions facilitate proactive learning or merely promote reactive corrections? We address these questions by examining major drug-to-drug interaction (DDI) errors in prescribing, a critical patient safety issue causing significant harm. Using a difference-in-differences model, we analyze data from a randomized field experiment involving 2.81 million prescriptions by ~1,700 doctors in India. Doctors were randomized into a treatment group receiving non-mandatory real-time informational interventions about potential errors or a control group without interventions. Results show that the treatment group prescribed ~8.6% fewer errors compared to controls, potentially reducing hospitalization costs by US\$4.8 million and saving an estimated 134 lives annually. Interestingly, doctors exposed to the intervention appear to effectively learn from the interventions, resulting in proactive avoidance of prescribing errors rather than mere corrections. Initially, error reduction is driven by immediate corrections, but sustained improvements result from proactive avoidance facilitated by learning. The effect is more pronounced among doctors with higher workloads. We find that the effectiveness of interventions remains consistent and does not negatively impact productivity or care quality. Managers can strategically utilize heterogeneity in treatment effects to target specific doctors and address errors more effectively. The feasibility of proactive learning reduces the reliance on continuous digital support, particularly benefiting healthcare systems in regions undergoing digitalization with limited internet connectivity.

MC34

Building B Level 2 B208

Distributionally Robust Optimization and Machine Learning

Invited Session

OPT: Machine Learning

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

Co-Chair: Siyuan Chen, Georgia Institute of Technology, Atlanta, GA, 30332, United States

1 - Data-Driven Conditional Optimization

Jonathan Li, Telfer School of Management, University of Ottawa, Ottawa, ON, Canada, Qinyu Wu

This talk introduces a distributionally robust framework for conditional optimization, aimed at addressing emerging challenges such as data scarcity, bias, and distribution shifts—issues commonly encountered in modern machine learning applications.

2 - Towards Learning An Optimal Robust Assortment

Miao Lu, Stanford University, Stanford, CA, United States, Yuxuan Han, Han Zhong, Zhengyuan Zhou, Jose Blanchet

Assortment optimization is a fundamental challenge in modern retail and recommendation systems, where the goal is to select a subset of products that maximizes expected revenue under complex customer choice behaviors. While recent advances in data-driven methods have leveraged historical data to learn and optimize assortments, these approaches typically rely on strong assumptions -- namely, the stability of customer preferences and the correctness of underlying choice models. However, such assumptions frequently break in real-world scenarios due to preference shifts and model misspecification, leading to poor generalization and revenue loss. Motivated by this limitation, we propose a robust framework for data-driven assortment optimization that accounts for potential distributional shifts in customer choice behavior. Our approach models uncertainty around a nominal choice model that generates observational data and seeks to maximize worst-case expected revenue. We first establish the computational tractability of robust assortment planning when the nominal model is known, then advance to the data-driven setting, where we design statistically optimal algorithms that minimize data requirements while maintaining robustness. Our theoretical analysis provides both upper bounds and matching lower bounds on the sample complexity, offering theoretical guarantees for robust generalization. Notably, we uncover and identify the notion of "robust item-wise coverage" as the minimal data requirement to enable sample-efficient robust assortment learning. Our work bridges the gap between robustness and statistical efficiency in assortment learning, contributing new insights and tools for reliable decision-making under uncertainty.

3 - Mixed-Feature Logistic Regression Robust to Distribution Shifts

Qingshi Sun, University of Southern California, Los Angeles, CA, United States, Nathan Justin, Andres Gomez, Phebe Vayanos

Logistic regression models are widely used in the social and behavioral sciences and in high-stakes domains, due to their simplicity and interpretability properties. At the same time, such domains are permeated by distribution shifts, where the distribution generating the data changes between training and deployment. In this paper, we study a distributionally robust logistic regression problem that seeks the model that will perform best against adversarial realizations of the data distribution drawn from a suitably constructed Wasserstein ambiguity set. Our model and solution approach differ from prior work in that we can capture settings where the likelihood of distribution shifts can vary across features, significantly broadening the applicability of our model relative to the state-of-the-art. We propose a graph-based solution approach that can be integrated into off-the-shelf optimization solvers. We evaluate the performance of our model and algorithms on numerous publicly available datasets. Our solution achieves a 408x speed-up relative to the state-of-the-art. Additionally, compared to the state-of-the-art, our model reduces average calibration error by up to 36.19% and worst-case calibration error by up to 41.70%, while increasing the average area under the ROC curve (AUC) by up to 18.02% and worst-case AUC by up to 48.37%.

4 - Finite-Adaptability via Explicit Partition

Zolykha Rezaei, Texas Tech University, Lubbock, TX, United States, Ningji Wei, Eojin Han

In many practical scenarios—especially those involving sequential decision making—professionals prefer policies that are not only interpretable but also straightforward to implement. In this research, we explore the concept of finite adaptability to construct robust optimization policies under uncertainty. More specifically, we address a general setting where uncertainty affects both the objective function and the constraints within the model. Our aim is to construct policies by partitioning the uncertainty set and assigning a contingent decision to each segment.

To solve the resulting robust optimization problem, we propose both exact and approximation methods. The exact methods include Benders' decomposition and dualization techniques for linear models. To enhance scalability and effective partition learning, we also develop approximation methods based on machine learning techniques such as support vector machines (SVMs) and decision trees.

5 - A Single-Loop First-Order Method for Robust Optimization

Manru Zong, The University of Hong Kong, Hong Kong, Hong Kong

Robust optimization is a powerful paradigm for decision-making under uncertainty. However, existing algorithms for robust optimization do not scale well, hindering its applications to modern large-scale decision problems. To address this challenge, we adopt the max-min-max framework recently proposed by Tu, Chen and Yue, and devise a single-loop first-order method, which is particularly suitable for large-scale robust optimization problems. We prove that our method enjoys a convergence rate of $O(K^{-0.5})$ for general non-smooth convex objective and constraint functions. We also prove that the rate can be improved to $O(K^{-1})$ under an additional smoothness or strong convexity assumption. Numerical experiments show that our method achieves a superior performance against existing approaches.

6 - Distributionally Robust Optimization Through the Lens of Submodularity

Karthik Natarajan, Singapore University of Technology and Design, Singapore, Singapore, Arjun Ramachandra, Divya Padmanabhan

We introduce a submodular ambiguity set and showcase its expressive power in modeling both discrete and continuous uncertainty. With discrete uncertainty, we show that a class of distributionally robust optimization problems is solvable in polynomial time and with continuous uncertainty, we show that it is solvable approximately up to an additive error in pseudo-polynomial time. In special cases, these problems can be solved as compact linear and semidefinite programs. We highlight the usefulness of these ambiguity sets particularly in optimization problems involving fairness criterion.

MC35

Building B Level 2 B209

Translational OR: Bridging Models and Societal Needs

Invited Session

Public Sector OR

Chair: Himadri Sen Gupta, Colorado State University Pueblo, 202 W Boyd St, Pueblo, CO, 73019, United States

Co-Chair: Andrés González, University of Oklahoma, Norman, OK, United States

1 - Raising the Rainbow Flag: Impact of Lgbtq+ Safety Cues on Service Provider Selection**Priyank Arora, University of South Carolina, Columbia, SC, United States, Mert Cetin, Dorothee Honhon**

In practice, several service systems are increasingly encouraging the use of LGBTQ+ safety cues from service providers to clients—for example, the use of rainbow flag stickers next to the service providers' names or pictures. As such, LGBTQ+ safety cues from service providers can help convey (1) LGBTQ+ Allyship: the service provider is an ally of the LGBTQ+ community, demonstrating support for LGBTQ+ rights and equality, and/or (2) Safe and Inclusive Space: the service provider is committed to providing a welcoming, non-judgmental, and affirming environment for LGBTQ+ clients. Intrigued by this emerging trend and inspired by calls to study the operations of inclusive service systems, we undertake this study to understand the effect of LGBTQ+ safety cues on the service provider selection process of clients, both from the LGBTQ+ community and the non-LGBTQ+ community. Specifically, our study aims to investigate whether such an LGBTQ+ safety cue given by the service provider is viewed positively or negatively by clients in their selection process. We also investigate how the effect of this cue on the selection process varies by the characteristics of the clients and the attributes of the service providers. Our findings provide insights into the decision-making process of the LGBTQ+ community, which can help organizations (re)design their operational practices and processes to improve equity and inclusion.

2 - Mitigation Planning for Community Resilience: a Two-Stage Robust Optimization Approach**Tasnim Faiz, University of Maryland, College Park, MD, United States, Kenneth Harrison**

Following high-impact hazard events, communities experience failures in their built environment, e.g., critical infrastructure and buildings. Damaged residential buildings cause injuries, fatalities, displacements, and high recovery costs, whereas damaged critical infrastructure reduces access to lifeline services. Proactive mitigation actions for the built environment can help avoid the adverse effects of hazards, reduce the risk associated with rare high-impact events, and improve community resilience against hazards. However, the complex interdependencies among a community's infrastructures and social functions, different resilience metrics, limited budget, and uncertainty associated with hazard events make mitigation planning challenging. A two-stage robust optimization framework is developed to address the uncertainties and the risk associated with rare high-impact events. Hazard uncertainty is defined by an ambiguity set, which facilitates tuning the degree of risk aversion when constructing high-impact hazard scenarios. The framework iteratively generates the worst-case hazard scenario and updates the mitigation decisions to minimize the resulting recovery and societal costs. A case study is presented to illustrate the benefit of the framework in developing alternative mitigation strategies for different levels of risk aversion to improve community resilience.

3 - Fair Resource Allocation Under Uncertain Utility Perception**Samuel Rodriguez, University of Oklahoma, Norman, OK, United States, Andres Gonzalez Huertas, Daniel Cabrera, Camilo Gomez**

Fair resource allocation in networked systems—such as energy grids, water infrastructure, or interdependent infrastructure network systems—hinges on how users derive value from the resources they receive. Yet when user preferences are uncertain or only partially observed, standard fairness assumptions may fail, raising concerns about the robustness of equity-driven allocation strategies. This paper examines the impact of utility function uncertainty within the Proportional Fairness framework—a widely adopted scheme that balances efficiency and fairness through logarithmic utility aggregation. We study two representative utility models: linear functions with constant marginal returns and power functions that capture diminishing returns. While linear utilities are structurally stable under uncertainty, power utilities are not, revealing vulnerabilities in fairness outcomes. To address this, we propose a Two-Stage Stochastic Programming approach that integrates uncertainty directly into the allocation process without abandoning fairness as a guiding principle. We demonstrate the method through a case study of Shelby County, TN, modeling the interdependent gas, electric power, and water networks. While the physical infrastructure remains fixed, variation in perceived utility—shaped by social vulnerability—is introduced as a stochastic element. Our results show how uncertainty reshapes allocation behavior, especially under scarcity, and provide both theoretical clarity and computational tools for practitioners seeking resilient and equitable resource distribution in complex networked environments.

4 - Saving Lives: Data-Driven Optimization in Animal Shelters**Qiuxia Chen, The University of Texas at Dallas, Richardson, TX, United States, Elena Katok, Ernan Haruvy**

Each year, an estimated 390,000 dogs are euthanized in the United States due to resource constraints at animal shelters. This study aims to improve live release rates by enhancing operational efficiency through data-driven decision-making. Recognizing the heterogeneity in shelter dog populations—where different conditions warrant different optimal lengths of stay—we address a critical gap in existing queuing models, which typically assume homogeneous service populations. We propose a four-step methodological framework that integrates queuing theory, discriminant analysis, survival analysis, and optimization to tackle the central operational challenge facing shelters: maximizing the live release rate while adhering to capacity constraints. Our model is built on 12 years of intake and outcome data from a large municipal animal shelter, encompassing over 240,000 individual records. The data include detailed information on animal characteristics, shelter conditions, and outcomes. Implementation of our model in 2024 led to a 27% increase in the shelter's live release rate. We project that this improvement can translate to saving between 1,400 and 1,800 dogs annually. These findings demonstrate the potential of operational research to generate meaningful and scalable impact in animal welfare.

5 - Community Operational Research: a 50-Year Retrospective and Prospective Analysis**Michael Johnson, University of Massachusetts Boston, Boston, MA, United States, Maria Pache de Athayde, Namesh Killemsetty**

This paper examines the evolution of Community Operational Research (COR) over the past 50 years, analyzing its methodologies, contributions to practice, and current research opportunities. Using a systematic analysis, we review over 800 scholarly articles according to academic disciplines, application areas, analytical approaches, geographies, characteristics associated with community-engaged research and, where possible, author characteristics. We identify key developments in analytic tools areas such as problem structuring methods, optimization, stochastic modeling, and data analytics, and assess COR's real-world impact in fields like housing, transportation, and environmental sustainability. We highlight the increasing diversity of actors engaged in COR. We discuss contemporary trends in community-engaged operations research and analytics that may or may not be explicitly associated with COR. By analyzing open problems and future directions, we aim to bridge the gap between theoretical advances and the needs of underserved communities.

MC36

Building B Level 2 B210

SpORts I

Invited Session

spORts

Chair: Stephen Hill, Samford University, Birmingham, AL, United States

1 - The Impact of Experience on Managerial Stints: Head-Coach Firing Versus Retiring in The National Football League 1970 through 2023

Michael Lapre, Vanderbilt University, Nashville, TN, United States, Zoe Segal

We study the impact of success and failure experience on the duration of managerial stints for head coaches in the modern era of the National Football League (NFL). We distinguish between stints ending with the firing of a head coach versus a head coach choosing voluntary retirement. From 1970 through 2023, excluding interim head coaches who only coached the latter part of a season, our dataset includes 264 head coaches representing 370 managerial stints. Using survival-analysis methods, we find that experience and NFL-specific control variables impact firing versus retiring differently.

2 - Return on Investment in Collegiate Athletics

Mike Gordon, Virginia Tech, Blacksburg, VA, United States

This study examines the operational efficiency of collegiate athletic departments by analyzing comprehensive financial data from public NCAA institutions. We investigate revenue generation sources, such as ticket sales and student fees, alongside expense categories including athletic student aid, medical support, and severance payments to terminated coaching staff. We specifically address the financial implications of athletic department debt levels and the associated servicing costs. A central focus is the economic impact of football programs, considering their operational scale and revenue potential across competitive divisions. Our findings provide valuable insights for athletic administrators, university leadership, and policymakers seeking to optimize resource allocation and enhance financial sustainability in collegiate athletics.

3 - Surviving NFL Survivor

Michael Gorman, University of Dayton, Dayton, OH, United States

NFL Survivor pools are a popular gambling game in which participants pick a single game per week, and if successful, the participant survives, and if not, the entrant is out. We use a combination of prediction methods, risk reduction, optimization and future discounting to improve outcomes in the popular pool game. We test the prediction and selection models over 19 NFL seasons (2006-2024) and find improved prediction, coupled with risk avoidance and appropriate future discounting leads to improved expected survival rates. Because of the high upset rates inherent in NFL games, portfolio strategies among multiple entries are explored to achieve reliable performance.

4 - Impact of Format Changes to the Uefa Champions League

Stephen Hill, Samford University, Birmingham, AL, United States

In this work we explore the impact of the recent changes to the format of the UEFA Champions League. We examine the competitive balance implications and also consider alternatives.

MC37

Building B Level 2 B211

Recent Advancements in Optimization under Uncertainty with Theory and Applications

Invited Session

OPT: Optimization Under Uncertainty

Chair: Beste Basciftci, Tippie College of Business, University of Iowa, Iowa City, IA, United States

1 - Interpretable Vaccine Administration and Inventory Replenishment Policies via Smooth-In-Expectation Decision Rules

Merve Bodur, University of Edinburgh, Edinburgh, United Kingdom, Zoha Sherkat Masoumi, Jim Luedtke

The effective management of vaccine vials and inventory is imperative for ensuring widespread immunization coverage. We aim to address the challenges associated with this problem, including the need for interpretable policies. We propose a multistage stochastic programming model, which offers flexibility to accommodate various operational aspects such as patient queues, clinic early closure considerations, and the requirement to serve every patient until exhausting vaccine inventory. For developing interpretable policies, we employ smooth-in-expectation decision rules, a recently proposed approach for multistage stochastic programs with mixed-integer recourse decisions and very many stages (e.g., hundreds). Leveraging these decision rules, we formulate and optimize the vaccine administration policies via a problem-specific flowchart design, alongside vial ordering decisions, facilitating interpretability and adaptability in real-world healthcare settings. We implement a batch stochastic gradient descent algorithm to solve the optimization problem, where for the initialization step we use the solution of a two-stage stochastic programming model we propose as an approximation to the multistage problem. Through extensive numerical experiments, we demonstrate the efficiency of the proposed approach and highlight the efficacy of various policies, including those considering patient queues.

2 - Distributionally Robust Optimization with Multimodal Decision-Dependent Ambiguity Sets

Beste Basciftci, Tippie College of Business, University of Iowa, Iowa City, IA, United States, Xian Yu

We consider a two-stage distributionally robust optimization (DRO) model with multimodal uncertainty, where both the mode probabilities and uncertainty distributions could be affected by the first-stage decisions. To address this setting, we propose a generic framework by introducing a phi-divergence based ambiguity set to characterize the decision-dependent mode probabilities and further consider both moment-based and Wasserstein distance-based ambiguity sets to characterize the uncertainty distribution under each mode. We identify two special phi-divergence examples (variation distance and χ^2 -distance) and provide specific forms of decision dependence relationships under which we can derive tractable reformulations. Furthermore, we investigate the benefits of considering multimodality in a DRO model compared to a single-modal counterpart through an analytical analysis. Additionally, we develop a separation-based decomposition algorithm to solve the resulting multimodal decision-dependent DRO models with finite convergence and optimality guarantee under certain settings. We provide a detailed computational study over two example problem settings, the facility location problem and shipment planning problem with pricing, to illustrate our results, which demonstrate that omission of multimodality or decision-dependent uncertainties within DRO frameworks result in inadequately performing solutions with worse in-sample and out-of-sample performances under various settings. We further demonstrate the speed-ups obtained by the solution algorithm against the off-the-shelf solver over various instances.

3 - Distributionally Fair Two-Stage Stochastic Programming by Bilevel Optimization

Yutian He, University of Iowa, Iowa City, IA, United States, Beste Basciftci, Qihang Lin

Two-stage stochastic programming (TSSP) is a fundamental framework for decision-making under uncertainty, where a first-stage decision is made before uncertainty is realized, followed by scenario-dependent second-stage decisions. While most TSSP literature focuses on cost minimization, fairness considerations in decision-making have largely been overlooked. Recently, Ye et al (2024) studied a one-stage stochastic program subject to a distributional fairness constraint, but similar development under the two-stage setting is still unavailable. In this work, we propose two models of TSSP under distributional fairness objective: one where the first- and second-stage decision-makers collaborate to ensure fairness, and another where only the first-stage decision-maker wants to ensure fairness, while the second-stage decision-maker only aims at minimizing the cost. To solve these models, we approximate the expectations by sample average and then propose reformulations resulting in mixed integer convex programs under certain assumptions. We further develop methods for finding computationally efficient lower bounds and introduce algorithmic approaches leveraging difference-of-convex programming reformulations to solve our problems effectively over large-scale instances. We provide case studies over sample problem settings, illustrating the value of the proposed approach and benefits of the proposed reformulations and solution algorithms.

4 - Contextual Stochastic Optimization for Omnichannel Multi-Courier Order Fulfillment Under Delivery Time Uncertainty

Tinghan Ye, Georgia Institute of Technology, Atlanta, GA, United States, Sikai Cheng, Amira Hijazi, Pascal Van Hentenryck

The paper studies a large-scale order fulfillment problem for a leading e-commerce company in the United States. The challenge involves selecting fulfillment centers and shipping carriers with observational data only to efficiently process orders from a vast network of physical stores and warehouses. The company's current practice relies on heuristic rules that choose the cheapest fulfillment and shipping options for each unit, without considering opportunities for batching items or the reliability of carriers in meeting expected delivery dates. The paper develops a data-driven Contextual Stochastic Optimization (CSO) framework that integrates distributional forecasts of delivery time deviations with stochastic and robust order fulfillment optimization models. The framework optimizes the selection of fulfillment centers and carriers,

accounting for item consolidation and delivery time uncertainty. Validated on a real-world data set containing tens of thousands of products, each with hundreds to thousands of fulfillment options, the proposed CSO framework significantly enhances the accuracy of meeting customer-expected delivery dates compared to current practices. It provides a flexible balance between reducing fulfillment costs and managing delivery time deviation risks, emphasizing the importance of contextual information and distributional forecasts in order fulfillment. This is the first paper that studies the omnichannel multi-courier order fulfillment problem with delivery time uncertainty through the lens of contextual optimization, fusing machine learning and optimization. The results provide actionable insights for e-commerce companies and online retailers to enhance service quality and customer satisfaction through efficient order fulfillment strategies that accounts for delivery time uncertainty.

5 - Distributionally Robust Optimization for Chemotherapy Scheduling Under Asymmetric and Multi-Modal Uncertainty

Zoe Zhu, The Ohio State University, Columbus, OH, United States, Xian Yu, Yu-Li Huang

We consider a real-world chemotherapy scheduling problem, where we need to design different appointment durations that can accommodate various patient types. From Mayo Clinic's real data, we observe that most patients' appointment durations are asymmetric (e.g., shorter durations tend to have a longer right tail). To address this, we consider a two-stage distributionally robust optimization (DRO) model under asymmetric and multi-modal uncertainty, where the overall ambiguity set is a mixture of two components. In the first layer, we use a variation-distance-based ambiguity set to capture the ambiguity on the mode probabilities. In the second layer, we consider a moment-based ambiguity set on the distribution in each mode, which includes a second-order partitioned statistics (semivariance) to account for the asymmetry in the random treatment duration distribution. We derive a closed-form expression for the worst-case expected cost and propose an efficient algorithm that can obtain a near-optimal solution in a quick fashion. Through numerical experiments based on both synthetic dataset and Mayo Clinic's real data, we illustrate the effectiveness of our approach and the benefits obtained by incorporating asymmetry information into the DRO formulation.

MC38

Building B Level 2 B212

Network Optimization Algorithms for Societal Applications

Invited Session

OPT: Network Optimization

Chair: David Shmoys, Cornell University, Ithaca, NY, United States

1 - Electrification of Heavy-Duty Vehicles: Challenges and New Algorithmic Insights

Haripriya Pulyassary, Cornell University, Ithaca, NY, United States, David Shmoys, Manxi Wu

While the adoption of electric vehicles (EV) offers substantial environmental benefits, it also introduces new challenges in vehicle routing due to range anxiety, long charging times, and sparse charging infrastructure. In this work, we approach EV routing through the lens of network flow theory. We begin with the analogs of three classical network flow problems: the problem of optimally routing a single EV, the maximum EV flow problem, and the minimum-cost EV flow problem. While these problems are NP-hard in the general case, we identify conditions under which they can be solved in polynomial time, and give exact, polynomial-time algorithms for this setting.

We then consider fleet electrification in long-distance logistics under various recharging modes. To model these, we introduce a general flow-based formulation defined over a charge-and-time augmented network. Due to the density of this network, directly solving the full integer program is impractical. Instead, we propose a variant of dynamic discretization discovery that significantly improves computational time. We also apply these methods in an empirical study of an instance constructed from real-world data.

2 - Leximin Fair Kidney Matching with Match Success Probabilities

Claire Chang, Cornell University, Ithaca, NY, United States, David Shmoys

The kidney matching process is critical for improving patient survival, fair access, and overall transplant success. With donor kidneys in short supply and high demand, the focus in many kidney matching algorithms is to maximize the number of exchanges. However, this convention disadvantages hard-to-match pairs (possibly due to sensitized patients with high antibody levels), many of whom may compete for a desirable donor. Furthermore, different exchanges have different probabilities of success based on factors like compatibility between patient-donor pairs and match rejection rates. Pre-existing work proposes a randomized algorithm that constructs a lottery over maximum cardinality matchings in order to optimize for the welfare of individual patient-donor pairs—that is, the leximin probability that any pair is included in a matching. Leveraging the Edmonds-Gallai decomposition, we exactly characterize the maximin coverage probability and present a simpler method for computing it. We also extend prior work to account for success probability of matches, and construct a lottery over just maximum-weight maximum-cardinality maximum weight matchings. Finally, we describe the Pareto frontier between matchings maximizing weight and cardinality. To better balance between the utilitarian and maximin objectives, we examine not only the leximin fairness objective, but also Nash social welfare as our metric for optimal lotteries.

3 - Fair Division Among Couples and Small Groups

Hannane Yaghoubzade, Cornell University, Ithaca, NY, United States, Paul Goelz

We study the fair allocation of indivisible goods across groups of agents, where each agent fully enjoys all goods allocated to their group.

We focus on groups of two couples and other groups of small size.

For two couples, an EF1 allocation¹,---\,one in which all agents find their group's bundle no worse than the other group's, up to one good\,---\,always exists and can be found efficiently.

For three and more couples EF1 allocations need not exist.

Turning to proportionality, we show that, whenever groups have size at most k , a PROP k allocation exists and can be found efficiently.

In fact, our algorithm additionally guarantees (fractional) Pareto optimality, and PROP1 to the first agent in each group, PROP2 to the second, etc., for an arbitrary agent ordering.

In special cases, we show that there are $\text{propone}\{\}$ allocations for any number of couples.

4 - Collective Communication in Data Centers via Multicommodity Flow

David Shmoys, Cornell University, Ithaca, NY, United States

Data centers today depend on a suite of collective communication protocols to distribute computation and aggregate results over large networks of GPUs. These protocols are typically designed based on simple structures that seek to achieve optimal latency in communication, while making effective use of the bandwidth limitations of the network. We present an approach to the automated design of these collective communication protocols based on multicommodity flow formulations, which therefore allow the design of protocols tailored to the specifics of the underlying network architecture, and can both achieve provable optimal latency, as well as optimizing other important metrics related to overall bandwidth utilization.

5 - Accelerating Decomposition Methods for Multicommodity Flow with Applications to Data Center Routing

Ricky Shapley, Cornell University, Ithaca, NY, United States

Multicommodity flow formulations can be used to generate effective collective communication primitives in GPU networks commonly used for data centers. These formulations are often at a scale that pose significant challenges for their solution. We introduce a number of significant advances in techniques for their efficient solution, including a technique that exploits the inherent symmetries in the architecture of the GPU networks to accelerate typical Dantzig-Wolfe style decomposition methods, coupled with insights that allow LP formulations to yield step-based routing solutions more commonly associated with integer programming formulations.

MC39

Building B Level 2 B213

Applications of Convex Optimization to Nonconvex Problems (part II)

Invited Session

OPT: Linear and Conic Optimization

Chair: Benjamin Grimmer, Johns Hopkins University, Baltimore, MD, United States

Co-Chair: Alex Wang, Purdue University, West Lafayette, IN, United States

Co-Chair: Kevin Shu, California Institute of Technology, Pasadena, CA, United States

1 - A Subgame Perfect Gradient Method**Alex Wang, Purdue University, West Lafayette, IN, United States, Benjamin Grimmer, Kevin Shu**

Optimization algorithm design has historically been concerned with worst-case a priori convergence rates. That is, convergence rates that can be guaranteed at the outset, before any information is learned about the problem instance. In the setting of unconstrained smooth convex optimization, the Optimized Gradient Method (OGM) achieves the optimal such convergence rate and is minimax optimal. In this talk, we introduce a stronger notion of optimality that requires a method to give optimal dynamic guarantees by optimally exploiting any “non-adversarialness” in the first-order oracle’s reported information. We then present a new first-order method achieving this notion of optimality: the Subgame Perfect Gradient Method (SPGM). SPGM is a refinement of OGM where the update rule is dynamically computed at runtime and can be implemented with minimal computational and storage overhead in each iteration. This talk is based on joint work with Benjamin Grimmer and Kevin Shu.

2 - Towards the Semidefinite Representability of Continuous Quadratic Submodular Minimization with Applications To Moment Problems**Samuel Burer, University of Iowa, Iowa City, IA, United States, Karthik Natarajan**

We discuss progress towards approximating continuous quadratic submodular minimization with bounds using semidefinite programming. We investigate two related moment problems arising in distributionally robust optimization and the computation of covariance bounds.

3 - Lagrangian Dual with Zero Duality Gap that Admits Decomposition**Santanu Dey, ISyE Georgia Tech, Atlanta, GA, United States**

For non-convex optimization problems with block structures and coupling constraints, on dualizing the coupling constraints the resulting Lagrangian relaxation becomes decomposable into blocks. However, the resulting Lagrangian dual can have non-zero duality gap due to the inherent non-convexity. In this talk, we propose a reformulation of the problem by adding redundant constraints, such that the Lagrangian dual obtained by dualizing the coupling constraints and the redundant constraints has zero duality gap and is decomposable into blocks. Joint work with Diego Cifuentes, Jingye Xu.

4 - Performance Estimation for Hölder Smooth Functions**Aaron Zoll, Johns Hopkins University, Baltimore, MD, United States, Benjamin Grimmer**

The minimization of Hölder smooth convex functions provides a family of problems interpolating from smooth to nonsmooth Lipschitz. This has proven to be a fruitful model for the design of universal, adaptive algorithms.

In this talk, we extend recent advances in computer-assisted design for first-order optimization over structured functions—known as Performance Estimation Problems (PEP)—to apply to Hölder smooth functions. We derive new “interpolation theorems” for Hölder smooth functions, enabling the computation of the worst-case performance of any proposed fixed stepsize first-order method (up to a small constant factor, approximately 1.2). This new theory is based on characterizations of “approximate cocoercivity” for associated tolerance parameters ϵ . As an application of this new PEP machinery, we identify new best-known methods for Hölder smooth minimization (requiring careful selection of both stepsizes and tolerances ϵ), improving on prior worst-case guarantees by a constant factor.

MC40

Building B Level 2 B214

Control of Service Systems

Invited Session

Service Science

Chair: Opher Baron, University of Toronto/ SiMLQ, Toronto, ON, Canada

Co-Chair: Nasser Barjesteh, University of Toronto, Toronto, ON, Canada

1 - Funding the Real Deal: Dynamic Moral Hazard with Adverse Selection**Izak Duenyas, University of Michigan-Ann Arbor, Ann Arbor, MI, United States, Feng Tian, Feifan Zhang, Peng Sun**

We study dynamic contracts to incentivize agents with private effort and cost information to generate Poisson breakthroughs. The principal offers a contract menu, inducing self-selection. We characterize optimal contracts under two agent types, revealing simple linear and one-switch structures, and provide a computationally efficient method to construct the full menu.

2 - Right to Repair and Competition in Durable Goods Markets**Cungen Zhu, Tongji University, Shanghai, China, People's Republic of, Luyi Yang, Chen Jin**

The right-to-repair movement advocates for stricter legislation that would require manufacturers of durable goods to facilitate independent repair. It is commonly believed that stronger repair laws would promote market competition—thereby benefiting consumers—and reduce e-waste, thereby protecting the environment. This paper builds a durable goods duopoly model to examine the economic and environmental implications of right-to-repair legislation. We find that when manufacturer competition is intense owing to limited product differentiation, tightening repair laws can further intensify competition and increase consumer surplus. However, when manufacturer competition is mild in a market of highly differentiated products, tightening repair laws can soften competition and hurt consumers. This suggests a polarization effect of strengthening right-to-repair legislation: it may increase competition when it is already strong but dampen competition when it is weak. In the former case, consumers may rush to replace their used products with new ones, which shortens the effective product lifespan and generates more e-waste. Overall, right-to-repair legislation can create a win-win situation that benefits both consumers and the environment, but this intended outcome is achieved in an unintended way: consumers are better off not through more affordable repair but through cheaper products, whereas the reduced environmental impact does not stem from less e-waste but from the use of newer products. Strikingly, tightening repair laws can lead to a lose-lose outcome for both consumers and the environment when they result in widespread availability and use of repair, only to benefit manufacturers by helping them escape the prisoner’s dilemma and thus increase profit.

3 - Dynamic Resource Allocation in Urban Search and Rescue Operations**Huiyin Ouyang, The University of Hong Kong, Hong Kong, Hong Kong, Zheqi Zhang, Sukriye Nilay Argon, Serhan Ziya**

We address resource allocation for urban search and rescue during disasters, focusing on deploying limited assets like helicopters across regions with varying population densities, distances, and rescue deadlines. The problem is modeled as an optimization of waiting and unmet request costs in a multi-class clearing system with deadlines.

Using a Markov decision process for a single resource and two request classes, we derive an optimal switching-curve policy. This insight informs a state-dependent heuristic for multiple resources and classes, which demonstrates near-optimal performance in numerical studies. A simulation based on Hurricane Harvey in 2017 validates the approach, showcasing its effectiveness in improving disaster response.

4 - Optimal Pricing for Queueing Systems with Loss-Averse Customers

Hongfan(Kevin) Chen, The Chinese University of Hong Kong, Shatin, China, People's Republic of, Jian Cao, Junfei Huang, Yongjiang Guo

We consider a single-server queueing system with uncertain service rate, in which customers have either disappointment-based loss-averse (DLA) or regret-based loss-averse (RLA) behavior. We explore the implications of DLA and RLA behaviors in terms of customers' decisions, the system manager's pricing strategies and other system metrics. We further emphasize the differences in these implications. We respectively consider the state-independent and state-dependent models to capture customers' DLA behavior and RLA behavior. We build a queueing game to first analyze customers' equilibrium decisions, and then investigate the system manager's pricing strategies from the perspectives of revenue and social welfare maximization. Our main findings are as follows. Given that customers' DLA behavior discourages customers from joining the queue, common intuition suggests that the manager should reduce the price to attract more customers. However, we observe that the revenue-optimal price may actually rise as customers' DLA level increases. In contrast, we demonstrate that RLA behavior has no impact on the system manager's pricing strategies in the context of revenue maximization, despite that RLA customers experience dis-utilities. From a welfare maximization perspective, in the presence of DLA and RLA behaviors, we demonstrate that interestingly, there may be no efficiency loss when the system manager implements the revenue-maximization pricing decision. Compared to revenue maximization, the system manager should consistently raise prices to mitigate the effects of customers' loss aversion. In a queueing system, it is crucial for the system manager to accurately identify the customers' loss aversion behaviors when determining prices.

5 - Fleet Sizing and Rebalancing in Free-Floating Bike-Sharing Systems

Jun Luo, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Hang Guo, Jiheng Zhang

Free-floating bike-sharing systems, widely adopted by operators such as Hellobike and Meituan Bike due to their great flexibility to users, present unique challenges in fleet sizing and rebalancing due to the spatial distribution of bicycles affecting user pickup behavior. We model these systems as load-dependent closed queueing networks, capturing spatial features through an explicit load-dependent form that reflects the relationship between bicycle density and user willingness-to-walk. Given the system's complexity, we apply fluid approximation techniques to develop a tractable optimization framework and propose two rebalancing policies: S-policy relocating bicycles to achieve a target system state, and Q-policy relocating a predetermined number depending on availability of bicycles from congested to shortage regions. Both policies are not only practical for real-world implementation but also asymptotic optimal under a large-market scaling regime. Our extensive simulation experiments demonstrate the superior performance of these policies compared to alternative approaches, confirming their effectiveness in maintaining appropriate bicycle distribution while minimizing operational costs across various system configurations.

MC41

Building B Level 3 B301

A Conversation on Health Equity Research in 2025

Panel Session

Health Applications Society

Co-Chair: Karen Hicklin, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, United States

1 - Moderator Panelist

Jennifer Lobo, University of Virginia, Charlottesville, VA, United States

2 - Moderator Panelist

Karen Hicklin, karen.hicklin@mountsinai.org, New York, NY, United States

We will gather experts in the area of health equity and operations research for a discussion regarding the current state of the field and emerging opportunities. The session will include an overview of the science of health equity research and social determinants of health, outline of funding and collaboration opportunities and, and perspectives on bringing an equity lens to all work we do within operations research and management science.

3 - Panelist

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

4 - Panelist

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

5 - Panelist

Jovan Julien, Georgia Institute of Technology, Atlanta, GA, United States

6 - Panelist

Chaitra Gopalappa, University of Massachusetts, Amherst, Amherst, MA, United States

MC42

Building B Level 3 B302

Advice for Early Career HAS Members

Panel Session

Health Applications Society

Co-Chair: Qiushi Chen, The Pennsylvania State University, University Park, PA, United States

1 - Moderator Panelist

Qiushi Chen, Pennsylvania State University, University Park, PA, United States

We propose a panel aimed at offering valuable insights to early career members of the Health Applications Society (HAS). In this panel, we plan to discuss several critical aspects of career development, including navigating funding opportunities beyond NSF, strategies for establishing a robust publication pipeline that balances Operations Research (OR) and interdisciplinary publications, student advising, and other challenges faced by early career researchers in the health modeling area.

2 - Panelist

Chaitra Gopalappa, University of Massachusetts, Amherst, Amherst, MA, United States

3 - Panelist

Oguzhan Alagoz, University of Wisconsin-Madison, Madison, WI, United States

4 - Panelist

David Hutton, University of Michigan, Ann Arbor, MI, United States

5 - Panelist

Pengyi Shi, Purdue University, West Lafayette, IN, United States

MC43

Building B Level 3 B303

Causal Inference for Healthcare Applications

Invited Session

Health Applications Society

Chair: Gabriel Zayas-Caban, University of Wisconsin-Madison, Madison, WI, United States

1 - Quasi-Experimental Designs for Learning Health Systems

Gabriel Zayas-Caban, University of Wisconsin Madison, Madison, WI, United States

This project aims to improve the way healthcare systems evaluate the use of risk prediction models, which are tools used to identify patients who might benefit most from specific medical interventions. Traditionally, new interventions are tested through randomized controlled trials, but these trials can be challenging to run due to their complexity and high resource demands. We propose using an alternative approach known as regression discontinuity (RD) designs. The main challenge is to develop new RD designs that allow for updates to models and interventions as new information and clinical insights become available.

2 - Wild Regenerative Block Bootstrap for Harris Recurrent Markov Chains

Kyuseong Choi, Cornell Tech, New York, NY, United States

We consider Gaussian and bootstrap approximation for the supremum of additive functionals of aperiodic Harris recurrent Markov chains. The supremum is taken over a function class that may depend on the sample size, which allows for non-Donsker settings; that is, the empirical process need not have a weak limit in the space of bounded functions.

We first establish a non-asymptotic Gaussian approximation error, which holds at rates comparable to those for sums of high-dimensional independent or one-dependent vectors. Key to our derivation is the Nummelin splitting technique, which enables us to decompose the chain into either independent or one-dependent random blocks. Additionally, building upon the Nummelin splitting, we propose a Gaussian multiplier bootstrap for practical inference and establish its finite-sample guarantees in the strongly aperiodic case. Finally, we apply our bootstrap to construct a uniform confidence band for an invariant density within a certain class of diffusion processes.

3 - When Additive Noise Meets Unobserved Mediators: Bivariate Denoising Diffusion for Causal Discovery

Dominik Meier, Cornell Tech, New York City, NY, United States, Sujai Hiremath, Promit Ghosal, Kyra Gan

Distinguishing cause and effect from bivariate observational data is a foundational problem in many disciplines, but challenging without additional assumptions. Additive noise models (ANMs) are widely used to enable sample-efficient bivariate causal discovery.

However, conventional ANM-based methods fail when unobserved mediators corrupt the causal relationship between variables. This paper makes three key contributions: first, we rigorously characterize why standard ANM approaches

break down in the presence of unmeasured mediators. Second, we demonstrate that prior solutions for hidden mediation are brittle in finite sample settings, limiting their practical utility. To address these gaps, we propose Bivariate Denoising Diffusion (BiDD) for causal discovery, a method designed to handle latent noise introduced by unmeasured mediators. Unlike prior methods that infer directionality through mean squared error loss comparisons, our approach introduces a novel independence test statistic: during the noising and denoising processes for each variable, we condition on the other variable as input and evaluate the independence of the predicted noise relative to this input. We prove asymptotic consistency of BiDD under the ANM, and conjecture that it performs well under hidden mediation.

Experiments on synthetic and real-world data demonstrate consistent performance, outperforming existing methods in mediator-corrupted settings while maintaining strong performance in mediator-free settings.

4 - Two-Way Fixed Effects Models with Network Interference

Mayleen Cortez-Rodriguez, Cornell University, Ithaca, NY, United States, Haiyun He, Christina Yu, Matthew Eichhorn

Accurate and precise estimation of causal effects from panel data is important to decision-makers in many settings, including healthcare, online platforms, and policy-making. Two-way fixed effects (TWFE) models and linear regression are popular tools for estimating causal effects from panel data, as they can account for outcomes or treatment effects that change over time and vary across individuals. While often criticized for relying on strong, ultimately untestable parametric assumptions, these models are simple to understand, and the literature offers both theoretical and practical guarantees in many settings. Classic causal inference methods rely on the “no interference” assumption, which states that each individual’s outcome depends only on their own treatment and not on the treatment of others. This assumption is violated in many real-world scenarios, such as the effect of vaccination on disease reduction due to herd immunity. Although there is a growing body of literature on causal inference under interference for non-panel data, relatively few results exist for the panel data setting. To address this, we analytically and empirically explore TWFE-inspired models that incorporate network interference at varying levels of heterogeneity. We compare the performance of linear regression-based estimators under different interference models and discuss conditions for model selection based on computational experiments.

MC44

Building B Level 3 B304

Generative AI for Decision-making

Invited Session

Decision Analysis Society

Chair: Guanting Chen, UNC Chapel Hill, Chapel Hill, NC, United States

1 - Risk Profiling and Modulation for LLMs

Yikai Wang, UNC Chapel Hill, Chapel Hill, NC, United States, Xiaocheng Li

Large language models (LLMs) are increasingly used for decision-making tasks under uncertainty; however, their risk profiles and how they are influenced by prompting and alignment methods remain underexplored. Existing studies have primarily examined personality prompting or multi-agent interactions, leaving open the question of how post-training influences the risk behavior of LLMs. In this work, we propose a new pipeline for eliciting, steering, and modulating LLMs' risk profiles, drawing on tools from behavioral economics and finance. Using utility-theoretic models, we compare pre-trained, instruction-tuned, and RLHF-aligned LLMs, and find that while instruction-tuned models exhibit behaviors consistent with some standard utility formulations, pre-trained and RLHF-aligned models deviate more from any utility models fitted. We further evaluate modulation strategies, including prompt engineering, in-context learning, and post-training, and show that post-training provides the most stable and effective modulation of risk preference. Our findings provide insights into the risk profiles of different classes and stages of LLMs and demonstrate how post-training modulates these profiles, laying the groundwork for future research on behavioral alignment and risk-aware LLM design.

2 - Why Agentic Design is Necessary for Data Analytics: The Limits of Multi-task Standalone LLM Inference

Daohong Tu, Stanford University, Stanford, CA, United States, Huaiyang Zhong, Guanting Chen

Agentic Large Language Models (LLMs) have gained attention for their ability to handle complex tasks through tool use, reasoning, and adaptive decision-making. However, the need for agentic design, beyond standard non-agentic approaches, has not been clearly established. In this work, we provide theoretical support for why agentic LLMs are necessary in data analytics. We demonstrate that, without agentic or reasoning-based mechanisms, using LLMs in standalone inference mode, even after pre-training and post-training enhancements such as supervised fine-tuning (SFT), or reinforcement learning from human feedback (RLHF), cannot reliably perform tasks in multi-task data science settings. Moreover, we theoretically and empirically show that mitigating this shortcoming would require modifications in the pre-training stage, which diverges from the standard LLM training paradigm. Our results characterize these limitations of standalone LLM inference and clarify where these approaches fall short in multi-task data science applications.

3 - Interactive LLM Agents for Data Science and Decision Analytics

Guanting Chen, UNC Chapel Hill, Chapel Hill, NC, United States, Yu Pan, Hanzhao Wang, Xiaocheng Li

We propose an interactive LLM agent framework designed specifically for data science and decision analytics. Unlike systems that focus primarily on automating computation, the agent emphasizes a professional yet interactive planning stage as its first step. Given a dataset, the agent collaborates with human users to interpret the data, clarify objectives, and determine an appropriate analysis or decision-making approach. This process reduces mistakes and hallucinations while ensuring that the chosen plan reflects domain-specific expertise contributed by practitioners, researchers, or instructors. Once consensus on the analysis plan is reached, the agent transitions from planning to execution by calling customizable local tool functions for computation and analysis. This design not only improves reliability by grounding results in trusted tools, but also makes the framework highly adaptable across different fields and problem domains. Through this combination of human expertise, structured planning, and tool integration, the framework provides a versatile pathway for incorporating LLM agents into professional workflows in data science, operations research, and decision analytics.

4 - Performative Risk Control for Reliable Decision-making by Generative AI

Zhun Deng, UNC at Chapel Hill, Chapel Hill, NC, United States, Victor Li, Qi Lei

Performativity of predictions refers to the phenomenon where prediction-informed decisions influence the very outcomes they aim to predict. This feedback loop is pervasive in both machine learning and decision-making. In this talk, we will introduce Performative Risk Control, a framework for calibrating predictive any blackbox models (e.g., LLMs) to achieve rigorous risk guarantees under performativity. We propose an iterative calibration process that ensures predictions are progressively improved while maintaining risk control, supported by theoretical guarantees. We also investigate the role of different risk measures and tail bounds in shaping the reliability of decision-making under performativity. Our study highlights new challenges and opportunities in decision-making theory with data feedback loop at the era of AI.

MC45

Building B Level 3 B305

Advances in Empirical and Field Research on Healthcare Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Zahra Mobini, Georgia Institute of Technology, Atlanta, United States

1 - Adverse Events Exposed: A Nationwide Study of Pediatric Emergency Departments

Mohamad Soltani, University of Alberta, Edmonton, AB, Canada, Mona Imanpoor Yourdshahy, Steven Shechter, Mahesh Nagarajan, Quynh Doan

In this presentation, we use data from pediatric emergency departments (EDs) across Canada to explore the impact of long ED waiting times on post-ED adverse events.

2 - Scaling up Timely Care: Combining Informational and Financial Interventions to Improve Follow-up Rates for High-Risk Urgent Care

Nazli Sonmez, ESMT GmbH, Berlin, Germany, Kamalini Ramdas, Manavi Sindal, Rengaraj Venkatesh, Priyanka Mahendrakar

Achieving impact at scale requires careful use of expensive resources. In this paper, we investigate the impact of combining a low-cost information intervention and a costly financial intervention to increase timely care access, in the context of emergency, high-consequence retina surgeries. In a two-stage randomized controlled trial of 2,100 emergency retina care patients conducted from 2019 to 2024 at Aravind Eye Hospital Pondicherry, all participants were first randomly assigned to view videos containing either emotional or technical content about the importance of timely follow-up. Next, financial incentives to follow-up were randomly provided among late patients. Notably, emotional videos increased on-time arrivals for the first appointment by 4.49%, demonstrating their potential for enhancing patient timely follow-up; technical videos had no significant effect on timely patient show-up rate. Moreover, financial incentives proved very effective for educated patients, increasing timely follow-up rates by 12–13%. However, these incentives had the opposite effect on less-educated patients, decreasing follow-up rates by 14–19%. This evidence sheds light on the importance of tailored interventions to improve the timely show-up rate of patients requiring high-consequence treatment to prevent needless blindness.

3 - Healthcare Algorithms and Disparities: Realized Value or Lost Opportunity?

Jiayi Liu, Virginia Tech, Blacksburg, VA, United States, Qi Wang, Ruba Aljafari, Anandhi Bharadwaj

The increasing adoption of artificial intelligence (AI) in healthcare holds promise for improving care quality, access, and patient outcomes. Yet concerns about algorithmic bias persist, particularly regarding its potential to exacerbate existing health disparities. This study examines the impact of healthcare algorithm adoption on racial disparities in clinical outcomes. Using six years of inpatient data analyzed through a difference-in-differences approach, we show that algorithm adoption improves outcomes for targeted groups, highlighting their potential for identifying and managing high-risk cases. However, these improvements are observed only among white patients, with no significant effects for non-white patients. Further analysis identifies substantial disparities in care delivery: non-white patients are much less likely to receive timely and intensive care following algorithm implementation—a key mechanism contributing to these disparities. These findings highlight the urgent need for greater scrutiny and reform in algorithm design and implementation to ensure equitable healthcare delivery.

4 - Beyond the Budget: the Unseen Costs of Maryland's Hospital Payment Reforms

Zahra Mobini, Georgia Institute of Technology, Atlanta, GA, United States, John McCandlish, Turgay Ayer

To rein in healthcare costs, Maryland implemented a global budget revenue (GBR) program for its hospitals, establishing yearly cost growth targets for all inpatient and outpatient services. Leveraging an extensive claims dataset for commercially insured individuals, our study empirically examines how the GBR influences care delivery patterns within hospitals.

5 - Pushing Past Patient Ratios: How a Recalculation of Workload Reshapes Nursing Imbalance and Alters Patient Care

Harshita Kajaria-Montag, Indiana University, Bloomington, IN, United States, RJ Niewoehner

Nurses play a pivotal role in patient care, but only if systems can hire enough of them and convince them to stay. Nurses make up the largest segment of the healthcare workforce, yet we find ourselves in the middle of a nursing crisis, and a shortage of nurses inevitably leads to overworked nurses. As a proxy for workload, most health systems carefully monitor the "nurse-to-patient ratio". We propose a fundamentally different tactic: using a painstakingly detailed workload tool, we observe how managers react to more accurate definitions of workload, including how they can better assign and balance work among their staff. We use archival data both before and after the launch of the workload tool from 67 unique nursing units operated by a large midwestern academic health system in the US. We observe over 4,000 nurses associated with over 60,000 patient encounters from January 2024 to March 2025. We find the introduction of the workload tool leads to load balancing, lower nurse attrition, and improved patient outcomes. Despite their virtues, most operations studies lack discussion of what actually happens when managers intervene and adjust server load. In contrast, we observe how managers deal with real-time fluctuations in load, enabling our study to push past generic suggestions of workflow towards actionable insights.

MC46

Building B Level 3 B306

Wisdom of the Crowd

Invited Session

Decision Analysis Society

Chair: Long Zhao, NUS DAO, Singapore, Singapore

1 - Enhancing Forecast Aggregation with Side Information About Experts**Fuqing Li, National University of Singapore, Singapore, China, People's Republic of**

Traditional forecast aggregation methods, such as the linear opinion pool, rely primarily on limited historical forecast errors to assign aggregation weights. However, in many real-world settings, additional background information about the forecasters—such as their professional training or area of expertise—is also available. This paper explores how such background information can be systematically incorporated to enhance the accuracy of aggregated forecasts. Using about 50,000 time series from the M4 Forecasting Competition dataset, we demonstrate that our proposed method often outperforms standard benchmark approaches including equally weighted aggregation, especially when expert background characteristics are indicative of individual forecasting performance.

2 - Weighted Average Ensemble with Regularization**Xiaohe Zhang, George Washington University, Washington, DC, United States, Han Su, Xiaojia Guo**

Aggregating forecasts from multiple experts often yields more accurate results than relying on a single expert. Among ensemble methods, the weighted average is known for its effectiveness and intuitive nature. In this paper, we propose a regularized ensemble method that integrates both historical performance and current forecasts of the experts to optimize weight selection. Our method learns the optimal weights by minimizing the variance of the ensemble forecast while incorporating a regularization term informed by historical performance. With a few popular choices of regularization, our proposed procedure also has a natural connection with the Bayesian framework. The appealing practical utility of the proposed ensemble method is demonstrated by empirical studies of the M5 competition and the Survey of Professional Forecasters.

3 - How Many Products to Pool in Forecast Aggregation? Maybe as Many as Possible**Long Zhao, NUS DAO, Singapore, Singapore**

Combining multiple forecasts into a consensus often yields high accuracy, with simple averaging frequently outperforming complex weighting schemes—a phenomenon termed the "forecast combination puzzle." Recent work by Chen & Zhao (2025) proposes pooled inference, leveraging forecasts of related variables to improve weighting. This complex method excels in the M4 dataset, even with large covariance matrices and limited data. These contradictory findings challenge the conventional wisdom, raising key questions: Why and when does pooled inference succeed? This paper addresses this gap, providing a theoretical and empirical analysis of the conditions under which the more products are pooled, the better the aggregation performance.

4 - Harnessing Collective Wisdom: Crowd Judging in Enhancing Dispute Resolution on Two-Sided Platforms**Chong Zhang, Tilburg University, Tilburg, Netherlands, Zhonghong Kuang**

Emerging digital platforms are introducing innovative mechanisms for resolving disputes by leveraging crowd judging as an alternative to traditional adjudication systems. This study investigates the effectiveness of crowd judging in improving dispute resolution on two-sided platforms, particularly in scenarios involving meticulous reviews and fastidious customers. We develop a game-theoretic model in which sellers can either present evidence proving their innocence or introduce misleading information to sway judges. By comparing conventional platform-based adjudication with the crowd judging mechanism, we aim to determine if the latter offers superior market efficiency and increased sellers' profits. Additionally, we explore the impact of information sharing among judges on the efficiency of the crowd judging process, assessing its potential to enhance decision-making accuracy.

MC47

Building B Level 3 B307

Optimization Topics VI

Contributed Session

Contributed

Chair: Xiu Yang, Lehigh University, Bethlehem, PA, 18015, United States

1 - Decentralized High-Dimensional Inference over Mesh Networks: a Unified Perspective**Marie Maros, Texas A&M University, College Station, TX, United States**

We consider the problem of solving high-dimensional statistical inference problems over a network of agents (with no coordinating agent) who have exclusive access to a fraction of the total available samples. In the high-dimensional setting, the problem dimension is much larger than the total number of available samples, making the problem ill-conditioned. Despite this, we empirically observe that obtaining a statistically meaning solution is possible with many existing decentralized schemes, given that underlying parameter to estimate lies in a low dimensional subspace. Our observations challenge the existing theories in two key ways: (i) most decentralized schemes do not break down as the problem dimensionality increases, and (ii) decentralized schemes that are expected to behave like one another behave very differently in high dimensions. To understand the behavior of decentralized optimization methods in high-dimensional inference we introduce a unified framework and analysis, allowing us to develop an understanding of the mechanisms enabling dimension independent performance of decentralized schemes.

2 - Quantum Circuit-Based Stochastic Optimization Integrated with a Quantum Probability Distribution Generator**Taihei Kuroiwa, The University of Electro-Communications, Chofu, Japan, Daiki Yamazaki, Yuya Ito, Kiria Yamashita, Daiki Sasaki, Ryosuke Koga, Keita Takahashi, Kodai Shiba, Tomah Sogabe**

Stochastic optimization is widely employed in domains such as energy system operation, supply-chain management, and financial portfolio optimization to account for uncertainty. However, as the number of scenarios grows, traditional scenario-based models and sample average approximation suffer from exponential increases in computational time and memory, limiting their applicability to large-scale problems.

In this work, we propose a quantum circuit-based stochastic optimization algorithm that integrates a pre-trained conditional quantum generative adversarial network (c-qGAN) directly into the quantum approximate optimization algorithm (QAOA). First, the conditional qGAN is trained offline to encode 2^n discrete scenarios into an n -qubit probability distribution generator. Next, the trained generator circuit is entangled with the QAOA circuit via controlled gates, enabling optimization decisions to incorporate uncertainty within the quantum circuit.

We apply our method to a renewable-integrated stochastic unit commitment problem, demonstrating that circuit depth and gate count scale polynomially with $\log_2$ (scenario count). Furthermore, on Qiskit simulators, our approach achieves solution quality on par with classical scenario-based optimization. Our results suggest that quantum circuit-based stochastic optimization could offer a viable, scalable alternative to classical methods as quantum hardware matures.

3 - Solving Constrained Marginally Convex Optimization Problems**Samuel Kwan, University of Chicago, Chicago, IL, United States, Siva Ramani, Andrew Schaefer**

In this paper, we attempt to solve constrained marginally convex optimization problems. We introduce an algorithm known as the upper bound coordinate descent. We show that this algorithm converges to a stationary point. We then apply this algorithm to solve an inverse Markov Decision Process problem, which is bilinear, hence marginally convex, by nature.

4 - Robust Admission via Two-Stage Stable Matching Under Ranking Uncertainty

Gustavo Angulo, Pontificia Universidad Católica de Chile, Santiago, Chile, Matías Giddings, Pablo Marshall

The Bachillerato Inicia UC program offers a pathway for students from Chilean technical high schools to articulate into undergraduate programs at Pontificia Universidad Católica de Chile. Upon applying, candidates rank up to three preferred programs. However, articulation is determined only after one year, based on their academic ranking within the cohort - a value unknown at the time of admission.

This setting gives rise to a two-stage admission problem with downstream matching constraints and exogenous uncertainty. The challenge is to select a feasible subset of students to admit, respecting the order of applicant scores, while guaranteeing that, under any realization of final rankings, each admitted student can be matched to one of their declared preferences, respecting program capacities.

We formalize this as a two-stage stable matching problem under ranking uncertainty and design an algorithm that characterizes the set of robustly admissible candidates. The model ensures articulation guarantees that are consistent with declared preferences and capacity constraints. We present implementation results from the 2024 and 2025 admission cycles. Joint work with Matías Giddings and Pablo Marshall.

5 - A Procedure to Assess Nonconvexity via Shortest Feasible Paths

Cameron Khanpour, Georgia Institute of Technology, Atlanta, GA, United States, Daniel Turizo, Samuel Talkington, Daniel Molzahn

We present a procedure that, given a compact non-convex set S , provides a metric representing how "close" S is to being a convex set. This metric is defined as the ratio of the *shortest feasible path* contained entirely within S to the Euclidean distance between the *endpoints* of such a path. Points from the set S are sampled uniformly through an efficient mixing algorithm for general QCQPs. By computing these ratios over all sampled pairs of points, we obtain an empirical distribution whose average provides a quantitative measure of the degree of convexity of S . We prove convergence of this sampling-based estimate to a true measure of convexity under mild assumptions. As an illustrative application, we assess the non-convex feasible sets of the optimal power flow (OPF) problem used to determine generator setpoints in electric power grids, independent of any specific objective function. We demonstrate the effectiveness of our procedure with numerical experiments using MATPOWER test cases.

6 - Resilient Consensus Without Finite TIME Detection Guarantees

Sarper Aydin, Harvard University, Cambridge, MA, United States, Orhan Eren Akgun, Stephanie Gil, Angelia Nedich

This paper investigates the asymptotic convergence of consensus processes under random malicious attacks with generalized time-varying attack probabilities. Agents leverage trust observations to identify their legitimate neighbors, but these generalized attack probabilities do not render any finite-time detection guarantees. In the absence of a finite-time detection, we demonstrate convergence of the consensus process in probability and expectation using misclassification probability bounds. We show that the weight matrices converge to the ideal weight matrix (which would have been used in the absence of malicious agents) in probability, and the consensus dynamics for legitimate agents reaches an agreement in probability and expectation. Numerical simulations validate our theoretical findings for several attack probability sequences.

7 - Distributionally Robust Optimization for Distribution Network Design Problem with Circular Transportation

Mianxing Zhong, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, China, Shenyang, China, People's Republic of, Defeng Sun, Lixin Tang

Distribution network design Problem, which involves facility location, warehousing, transportation decisions, is of fundamental importance for companies to be competitive on the market. In this paper, we study a three-level logistic distribution network problem with circular transportation and direct shipment under the uncertainty of discount factor on transportation. To describe the uncertainty, we introduce a Wasserstein distance-based ambiguity set and establish a two-stage distributionally robust optimization model which is converted into a certain mixed integer linear programming model. We propose a Benders decomposition approach to solve the problem which is enhanced through valid inequalities, combinatorial Benders' cuts and high density Pareto cuts. Computational experiments demonstrate that the performance of two-stage distributionally robust optimization model solutions can hedge against the uncertainty in discount factor on transportation flexibly. Finally, this paper shows the effectiveness and robustness of the model and provides managerial insights for distribution network design and management.

8 - Noise-Aware Optimization for Variational Quantum Algorithms

Xiu Yang, Lehigh University, Bethlehem, PA, United States

The success of variational quantum algorithms (VQAs) relies on classical optimization algorithms that allow function information distorted by sampling and hardware noise. We introduce a noise-aware optimization scheme that estimates gradient and curvature information from finite-difference samples then optimizes the objective using a quasi-Newton algorithm framework. Applied to the Quantum Approximate Optimization Algorithm (QAOA) for Max-Cut, the method aims to cut circuit-evaluation costs and maintain steady progress with noise. Preliminary numerical results demonstrate the superior performance compared to standard gradient-free approaches.

MC48

Building B Level 3 B308

Special Issue of Decision Analysis on AI

Invited Session

Decision Analysis Society

Chair: Ali Abbas, University of Southern California, Los Angeles, CA, United States

1 - Chatgpt vs. Experts: Can Genai Develop High-Quality Organizational and Policy Objectives?

Jay Simon, American University, Washington, DC, United States, Johannes Siebert

We explore the efficacy of GenAI for value-focused thinking, specifically the ability to generate high-quality sets of objectives for organizational or policy decisions. Overall, we find that most of the GenAI objectives are viable individually, but the sets as a whole are highly flawed in the given decision setting. They often include non-essential considerations while omitting important ones. In addition, they are redundant and non-decomposable, often because of a tendency to include means objectives even when explicitly instructed not to. Prompt improvements mitigate the problems only slightly. The results provide further evidence of the importance of a human in the loop; in this case, GenAI tools are helpful for brainstorming objectives, but an expert is needed before the results are used to support decision making. To facilitate this, a four-step approach incorporating the relative strengths of both GenAI and decision analysts is presented and demonstrated.

2 - Comparing Humans and LLMs Alternative, Preferences , and Uncertainties for Decisions

Ali Abbas, University of Southern California, Los Angeles, CA, United States, Barbara Mellers, Rachel Monger

This talk presents the results of two experiments that compare the relevance and redundancy of inputs to decision quality from individuals and Chat GPT. The experiments were conducted using the social media platform Ahoona. Individuals were asked to provide inputs to decision quality without seeing the GPT inputs. A second experiment compared the individual inputs to those of Chat GPT with regeneration and probing. The work also explores the impact of exposure to decision education on this comparison.

3 - Toward Synthesizing AI and Decision Analysis: Initial Thoughts

Jeffrey Keisler, University of Massachusetts-Boston, Boston, MA, United States

We consider qualitatively which DA and adjacent tasks are performed well by AI, by humans, or both—and how important these tasks are to the decision process. This lens suggests directions for synthesizing AI and DA into something more than the sum of their parts. In this, AI's greatest value may lie not in formal modeling, but in helping connect those models to reality.

4 - AI as Opportunity for Decision Education

Chris Spetzler, Decision Education Foundation, Portola Valley, CA, United States

AI holds great promise in augmenting human intelligence, specifically in the learning and teaching of decision skills. This presentation will share initial experiences using AI in decision education. Decision Education Foundation's (DEF) work with high school students at The Socratic Experience (TSE) online high school and at the Decision Quality Center at the Illinois Institute of Technology are two current examples with additional applications in store. At TSE, Eric Horvitz's Decision Insights VIII helped students evaluate potential ways to approach a gap year, while reinforcing prior learning on a highly relevant and relatively complicated application of decision analysis to students' lives. At Illinois Tech, an AI decision tool is being developed to allow students to coach an student/bot facing a challenging decision. This promising tool helps students to both see and demonstrate the value of their decision skills in an active way with immediate feedback. The DQ Center, in partnership with DEF, is working to ensure all Illinois Tech students learn principles of Decision Quality as a part of their undergraduate experience. The AI tool holds the promise of ensuring high and consistent standards across instructors for teaching, learning and assessment of Decision Quality.

MC49

Building B Level 3 B309

Practice Driven Research in Supply Chain

Invited Session

MSOM: Supply Chain

Chair: Liang (Leon) Xu, Singapore Management University, Singapore, Singapore

1 - Suppliers' Self-Disclosed Vulnerability and Brand Perception in Sustainable Supply Chain Practices

Amolwan Jittamai, University of Houston, Houston, TX, United States, Ravi Aron

This research examines how voluntary self-disclosure of negative information by a brand about its supply chain practices affects consumer perceptions of the brand. Using a mixed experimental design combining between-subjects (brand positioning and level of self-disclosed negative information) and within-subjects (whether the brand self-disclosed negative information) factors, we investigate how transparency through self-disclosure of negative information regarding the brand's sustainable supply chain practices influences consumers' level of trust, perceived product quality, purchase intention, and willingness to pay. In addition, we investigate whether a brand can game the market by strategically under-investing in sustainable practices and disclosing its deficits to benefit from consumers' trust.

We find that when comparing two brands that have implemented the same levels of sustainable practices, the brand that fully self-revealed its deficits is rewarded by consumers. Both trust and perceived product quality generated through self-disclosed negative information led to positive consumer outcomes regardless of brand positioning. In addition, we find that a brand cannot act strategically by under-investing in sustainable practices as consumers penalize suppliers with lower trust for under investment in sustainable practices even if the deficits are self-revealed. This research extends transparency studies by demonstrating that self-revealed vulnerability by disclosing negative information can build consumer trust and enhance perceived product quality. However, a brand cannot behave in an opportunistic and strategic manner and attempt to free ride on the economic goodwill generated by the self-revelation of deficits and under invest in sustainable practices.

2 - AI-Driven Bidding in Procurement Auctions

Jingyi Tian, University of Houston, Houston, TX, United States, Meng Li

As online auction platforms advance, bidders increasingly turn to artificial intelligence (AI) for strategic advantages. This paper studies AI-driven bidding behavior in second-price procurement auctions through conducting experiments with agents trained by a model-free, value-based reinforcement learning algorithm: the Deep Q-Network (DQN). Through extensive simulations under varying competition levels and randomized initializations, we identify a robust tendency toward overbidding as the algorithm converges to a stable state. Bids exceed costs consistently, and the gap widens as competition intensifies. This behavior departs from the incentive-compatible Nash equilibrium, resulting in inflated procurement costs. By connecting DQN's iterative value-based updating mechanism to the reward structure in repeated auction games, we show that this systematic overbidding arises due to the internalization of positive externalities created by elevated bids, which enhance competitors' expected rewards. This interdependent dynamic destabilizes truthful bidding, fostering implicit collusive behavior despite the absence of explicit coordination. To mitigate these collusive outcomes, we propose and experimentally evaluate two targeted strategies: limiting the number of bidders and imposing lower reserve-price ceilings. Supplemental experiments involving enriched informational inputs to the algorithm further reveal the nuanced ways in which the information environment influences the learning process. Our findings uncover a novel, unintended form of algorithmic collusion in procurement auctions, highlight critical gaps in existing antitrust policies, and provide practical design and regulatory insights to help auction designers embrace AI-driven bidding without compromising market competitiveness and efficiency.

3 - Automated Decision Framework for Multi-Paradigm Warehouse Slotting Optimization

Aakash Bhatt, Chewy, Inc., San Jose, CA, United States

Modern e-commerce fulfillment operations face growing complexity in storage location assignment across multi-paradigm facilities where automated robotic systems and manual pick modules coexist. Traditional approaches either focus on single storage paradigms or apply isolated heuristics to each system, leading to locally optimized decisions that undermine global fulfillment efficiency.

We present a novel two-stage mixed integer programming framework that fully automates warehouse slotting decisions across all storage paradigms. Its core innovation is a hierarchical decomposition approach that handles a 70K+ SKU catalog across 100K+ locations while maintaining computational traceability. The first-stage MIP incorporates cross-paradigm constraints—pick cost, machine capacity, and replenishment efficiency—to optimize macro-allocation between robotic and manual storage modules. The second-stage MIP then solves location-specific assignments with affinity-based travel minimization, achieving global optimality beyond the reach of traditional sequential methods. The resulting move instructions are automatically transmitted to the warehouse management system, enabling iterative convergence toward optimal item placement.

Thorough simulation testing and successful implementation at a Chewy fulfillment center (processing 100K+ daily orders) demonstrate improvements over both manual expert slotting and single-paradigm optimization: 15% increase in pick rates and 20% reduction in cycle times compared to previous best-practice configurations. The system re-optimizes weekly with adaptive implementation pacing (3-5% of inventory relocated per cycle) to balance disruption costs against performance gains. This autonomous approach represents an advancement in warehouse optimization by bridging theoretical models with implementation practicality in complex multi-paradigm environments.

MC50

Building B Level 3 B310

Advances in Healthcare Operations and Analytics

Invited Session

MSOM: Service Operations

Chair: Michael Lingzhi Li, Harvard Business School, Boston, MA, 02163, United States

1 - Routing with Delayed Revelation of Feature Information

Xiaoquan Gao, Singapore Management University, Singapore, Singapore, Pengyi Shi

We study the routing problem for heterogeneous customer arrivals, where individual features are initially unavailable and only revealed after evaluation. This delay can lead to inefficiencies in resource allocation, particularly in systems with diverse user classes requiring personalized assessments. To address this challenge, we propose a two-stage approach. In the first stage, individuals are allocated to different programs based on arrival counts, historical class data, and facility occupancy, enabling proactive resource planning. In the second stage, individualized placement decisions are made after evaluations. We establish a performance bound relative to a full-information Markov Decision Process (MDP), provided that second-stage routing follows treatment effect rankings. Computationally, we develop a fluid model and apply constrained linear quadratic regulation (LQR) to derive an asymptotically optimal, easy-to-implement linear decision rule for the first stage. Simulations calibrated with real-world data from community corrections demonstrate that our approach outperforms existing threshold-based policies, improving staffing and resource coordination without requiring immediate judicial decisions for each individual.

2 - Causal Reinforcement Learning and Centaurs: Making AI Impactful in Healthcare

Soroush Saghaian, Harvard University, Cambridge, MA, United States

There is increasing evidence that Machine Learning and Artificial intelligence algorithms can be used to enhance clinical care. In this talk, I address two critical aspects that can significantly improve the impact of such algorithms in healthcare practices: (1) moving beyond associations and creating algorithms capable of causal reasoning under ambiguity, and (2) a human-algorithm “centaur” model of care and decision-making, in which the power of human intuition is combined with the outstanding capabilities of algorithms. I describe our latest research on these subjects at the Public Impact Analytics Science Lab (PIAS-Lab) at Harvard, and discuss findings based on our various collaborations with the Mayo Clinic, Mass General Hospital, and some other public and private organizations.

3 - Holistic Optimization of Operating Rooms

Kimberly Villalobos Carballo, New York University, New York, MA, United States

Surgeries are a key component of hospitals’ revenue and a vital element of patient care. We propose a holistic framework for operating room (OR) optimization using robust and discrete optimization techniques. Our three-stage approach includes (1) demand forecasting, (2) tactical definition and allocation of OR blocks to surgeons, and (3) operational-level online surgery scheduling. Simulations demonstrate that our method can reduce inpatient census by 6–10% and increase OR utilization by 32%, highlighting its potential to enhance both care delivery and hospital efficiency.

4 - Exact Sensitivity Analysis of Markov Reward Processes via Algebraic Geometry: Application to Cost-Effectiveness

Timothy Chan, University of Toronto, Toronto, ON, Canada, Muhammad Maaz

We introduce a new approach for deterministic sensitivity analysis of Markov reward processes, commonly used in cost-effectiveness analyses, via reformulation into a polynomial system. Our approach leverages cylindrical algebraic decomposition (CAD), a technique arising from algebraic geometry that provides an exact description of all solutions to a polynomial system. While it is typically intractable to build a CAD for systems with more than a few variables, we show that a special class of polynomial systems, which includes the polynomials arising from Markov reward processes, can be analyzed much more tractably. We establish several theoretical results about such systems and develop a specialized algorithm to construct their CAD, which allows us to perform exact, multi-way sensitivity analysis for common health economic analyses. Finally, we apply it to two case studies, one with synthetic data and one that re-analyzes a previous cost-effectiveness analysis from the literature, demonstrating advantages of our approach over standard techniques. We develop an open-source software package that implements our algorithm. Our software and code are available at: <https://github.com/mmaaz-git/markovag>.

5 - Measuring the Heterogeneous Effect of Emergency Department Boarding on Inpatient Length of Stay: Toward Efficient and Equitable Inpatient Bed Assignment

Xiaole (Alyssa) Liu, New York University Stern School of Business, New York, NY, United States, Jing Dong, Martin Copenhaver, Yosef Berlyand

Emergency Department (ED) boarding, the practice of holding admitted patients in the ED while awaiting an inpatient bed, is a common challenge in hospitals worldwide. It is a major contributor to ED crowding and raises concerns about the quality of care. When inpatient bed capacity is limited, hospitals must decide how to prioritize boarding patients for admission in a way that is both efficient and equitable. In this study, we investigate the heterogeneous causal impact of ED boarding on inpatient length of stay (LOS), a key metric influencing downstream capacity. Using methods from double machine learning, instrumental variables, and heterogeneous treatment effect modeling, we identify patient subgroups for whom prolonged boarding has different effects on LOS. We then integrate these individualized effect estimates with queueing models and policy learning techniques to design inpatient bed assignment policies that reduce overall ED boarding time. A key feature of our approach is the explicit integration of fairness considerations into operational policy design. First, we construct policies that enforce group fairness. Second, we propose a class of policies that navigates the tradeoff between system efficiency and individual fairness. We validate our proposed policies using a high-fidelity, trace-driven simulation model calibrated with real hospital data. Results show that our policies can substantially reduce the average ED boarding times by over 75% compared to the current bed assignment policy, while maintaining strong fairness performance.

MC51

Building B Level 3 B311

LLMs for Decision-Making

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Linwei Xin, Cornell University, Ithaca, NY, United States

Co-Chair: Jingyuan Yang, University of Chicago, Booth School of Business, Chicago, IL, United States

1 - Large Language Models for Market Research: a Data-Augmentation Approach

Mengxin Selene Wang, The University of Texas at Dallas, Richardson, TX, United States, Dennis Zhang, Heng Zhang

Large Language Models (LLMs) have transformed artificial intelligence by excelling in complex natural language processing tasks. Their ability to generate human-like text has opened new possibilities for market research, particularly in conjoint analysis, where understanding consumer preferences is essential but often resource-intensive. Traditional survey-based methods face limitations in scalability and cost, making LLM-generated data a promising alternative. However, while LLMs have the potential to simulate real consumer behavior, recent studies highlight a significant gap between LLM-generated and human data, with biases introduced when substituting between the two. In this paper, we address this gap by proposing a novel data augmentation approach that efficiently integrates LLM-generated data with real data in conjoint analysis. Our method leverages transfer learning principles to debias the LLM-generated data using a small amount of human data. This results in statistically robust estimators with consistent and asymptotically normal properties, in contrast to naive approaches that simply substitute human data with LLM-generated data, which can exacerbate bias. We validate our framework through an empirical study on COVID-19 vaccine preferences, demonstrating its superior ability to reduce estimation error and save data and costs by 24.9% to 79.8%. In contrast, naive approaches fail to save data due to the inherent biases in LLM-generated data compared to human data. Another empirical study on sports car choices validates the robustness of our results. Our findings suggest that while LLM-generated data is not a direct substitute for human responses, it can serve as a valuable complement when used within a robust statistical framework.

2 - Auto-Formulating Dynamic Programming Problems with Large Language Models

Jingyuan Yang, University of Chicago, Booth School of Business, Chicago, IL, United States

Dynamic programming (DP) is a fundamental method in operations research, but formulating DP models has traditionally required expert knowledge of both the problem context and DP techniques. Large Language Models (LLMs) offer the potential to automate this process. However, DP problems pose unique challenges due to their inherently stochastic transitions and the limited availability of training data. These factors make it difficult to directly apply existing LLM-based models or frameworks developed for other optimization problems, such as linear or integer programming.

We introduce DP-Bench, the first benchmark covering a wide range of textbook-level DP problems to enable systematic evaluation. We present Dynamic Programming Language Model (DPLM), a 7B-parameter specialized model that achieves performance comparable to state-of-the-art LLMs like OpenAI's o1 and DeepSeek-R1, and surpasses them on hard problems. Central to DPLM's effectiveness is DualReflect, our novel synthetic data generation pipeline, designed to scale up training data from a limited set of initial examples. DualReflect combines forward generation for diversity and backward generation for reliability. Our results reveal a key insight: backward generation is favored in low-data regimes for its strong correctness guarantees, while forward generation, though lacking such guarantees, becomes increasingly valuable at scale for introducing diverse formulations. This trade-off highlights the complementary strengths of both approaches and the importance of combining them.

3 - From Qualification Exams to Practice: Testing LLMs on Solving Stochastic Modeling Problems

Tianyi Peng, Columbia University, New York, NY, United States, Akshit Kumar, Yuhang Wu, Assaf Zeevi

Large language models (LLMs) have exhibited capabilities comparable to those of human experts in various fields. However, their modeling abilities---the process of converting real-life problems (or their verbal descriptions) into sensible mathematical models---have been underexplored. In this work, we take the first step to evaluate LLMs' abilities to solve stochastic-modeling problems, a model class at the core of Operations Research (OR) and decision-making more broadly. We manually procure a representative set of graduate-level homework and doctoral qualification-exam problems and test LLMs' abilities to solve them. We further leverage SimOpt, an open-source library of simulation-optimization problems and solvers, to investigate LLMs' abilities to make real-world decisions. Our results show that, though state-of-the-art LLMs demonstrate proficiency on par with human experts in both classroom and practical settings, a nontrivial amount of work is still needed to reliably automate the stochastic-modeling pipeline in reality.

4 - Synthetic but Not Infinite: How Much LLM-Generated Data to use in Market Research

Linwei Xin, Cornell University, Ithaca, NY, United States, Qichuan Yin

Large Language Models (LLMs) have transformed data generation by enabling scalable synthetic data augmentation. This paper explores the optimal use of LLM-generated synthetic data in market research, with a focus on improving choice model estimation under limited training data. We introduce a framework, supported by a theoretical analysis, that identifies the optimal ratio of synthetic to real data and characterizes the resulting estimation error of this hybrid approach.

MC52

Building B Level 3 B312

Advances in Sustainable Operations

Invited Session

MSOM: Sustainable Operations

Chair: Zuguang Gao, University of California, Irvine, Irvine, CA, 92697, United States

1 - From Wheels to Wallet: Optimizing Transit Service, Pricing, and Funding Allocation

Jingwei Zhang, Cornell University, Ithaca, NY, United States, Owen Wu

Public transit systems in many U.S. cities face a dual crisis of declining ridership and financial instability, exacerbated by income segregation that creates diverse, often unmet, rider needs. This paper addresses the critical challenge of how transit agencies can optimize pricing, service levels (time-related factors), and the internal allocation of limited budgets across income-segregated regions to maximize total ridership. We develop a game-theoretic, bilevel optimization model where a budget-constrained transit agency sets differentiated or uniform fares and service times, anticipating the mode choices of utility-maximizing consumers from distinct high- and low-income regions.

Our analysis reveals that under a differentiated policy, the optimal allocation of the total budget between income regions is non-monotonic, depending critically on income disparity, service improvement costs, and overall budget availability. For instance, substantial budgets may favor high-income regions if service improvements are inexpensive and disparity is high, but under fiscal scarcity, low-income regions might receive more to prevent ridership collapse. We also demonstrate the potential for bidirectional cross-subsidization, challenging common assumptions. Comparing differentiated policies with uniform benchmarks, we identify conditions where differentiation not only improves total ridership but can also yield targeted benefits for specific income groups. These findings offer actionable insights for transit agencies, suggesting that strategically differentiating pricing, service, and particularly budget allocation based on income-related sensitivities can more effectively achieve ridership goals than traditional uniform policies, providing a novel operations management approach to a pressing public service challenge.

2 - Climate Change and Thermostat Behavior: Empirical Evidence, the Impact, and Interventions

Michael Blair, Wilfrid Laurier University, Waterloo, ON, Canada, Saed Alizamir, Shouqiang Wang

Using a rich micro-level dataset, we empirically analyze smart thermostat data to understand the relationship between households' thermostat settings and their ambient environment. Within this data we observe two key decision levers that households use to manage their home heating and cooling. Combining a variety of methodological tools including Dynamic Linear Models, random effects, and Bayesian Statistics, we develop models for these behavior levers and study how households are impacted by varying weather conditions. A critical insight from these models is that households typically react to the weather in a comfort-seeking manner, leading to increased energy usage. We also find that this behavior is more prevalent in households that routinely override the settings they have pre-programmed into the thermostat. With these insights in mind, we leverage well-established models to create realistic estimates of future weather conditions under a variety of climate change scenarios and combine these scenarios with our statistical models to estimate the impact of climate change on how households use their thermostats. Specifically, we predict future household behavior and investigate the resulting changes in energy usage for heating and cooling. These changes have a profound impact on the operations of the energy grid specifically relating to operating costs, grid capacity, and reliance on intermittent renewable energy sources. The insights derived from our research provide a valuable framework to evaluate and design interventions. Furthermore, we discuss two changes that smart thermostats manufacturers could make to potentially alleviate the behavioral challenges identified.

3 - Multi-Period Aggregation of Distributed Energy Resources

Lisa Hillas, University of Auckland, Auckland, New Zealand, Zuguang Gao

This paper looks at the problem of how distributed energy resources (DERs), such as household-level solar panels, should be integrated into the wholesale electricity market. We assume that DER owners have access to batteries that allow them to transfer power across time periods. We look at how DER owners should choose to consume, exchange with the market, and buy or sell power. Additionally, given the choices of the DER owners, we look at how an aggregator whose role is to procure electricity from DERs and sell the electricity in the wholesale market should set prices to the DER owners.

MC53

Building B Level 3 B313

Retail Analytics and Marketplace Innovation

Invited Session

Revenue Management

Chair: Hansheng Jiang, University of Toronto, Rotman School of Management, Toronto, ON, Canada

1 - Bias in Offline Retailing Experiment: Evidence and Solution

Jiayi Zhang, Washington University in St. Louis, St Louis, MO, United States, Zhiqi Zhang, Dennis Zhang, Ruohan Zhan

Product-based randomization, where a random subset of products receives treatment while others serve as controls, is commonly used in offline retailing experiments. We show this can result in biased causal inference due to consumer substitution and complementarity behaviors. We propose a double-machine-learning-based debiasing method with theoretical guarantees and empirical performance.

2 - Can Personalized Pricing be Progressive?

Titing Cui, University of Tulsa, Tulsa, OK, United States, Max Biggs, Michael Hamilton, Enfeng Xing

Traditional welfare analysis evaluates pricing policies based on aggregate revenue and customer surplus. However, modern approaches increasingly emphasize equity by asking not just how much aggregate utility is derived, but by whom. In this work, we examine market outcomes under a new notion of *proportional surplus*, which measures the percentage of a customer's valuation that is retained as surplus. Proportional surplus provides a more equitable lens by which to view market outcomes, assigning greater weight to utility gained by customers with the least ability to pay. Further, we demonstrate that although traditional customer surplus often decreases when firms price discriminate, proportional surplus increases under moderate levels of revenue-maximizing price discrimination in many natural markets. These findings challenge the conventional belief that price discrimination harms customers, and suggests instead that modest amounts of price discrimination can improve outcomes for both firms and the markets that they serve.

3 - Learning to Optimize Service Upgrades in Systems with Reusable Resources

Xuchen Liu, University of California, Berkeley, Berkeley, CA, United States, Huiwen Jia

We study a revenue management problem in a service system with reusable resources over a finite horizon T . Customers arrive according to a Poisson process and have exponentially distributed service durations, both with unknown parameters. The system operates under a fixed base price, while also offering an implicit upgrade option that allows customers to extend their service duration by a prespecified factor at a fixed discounted unit price. Customers choose between the base and upgraded options according to an unknown preference model. The objective is to learn the optimal upgrade option from a discrete set of candidates to maximize cumulative revenue. We propose a Batch Upper Confidence Bound (BUCB) algorithm that jointly (i) learns the system dynamics, i.e., the arrival rate, service rate, and upgrade choice probability for each option, and (ii) maximizes revenue under incomplete information. We establish high-probability bounds on parameter estimation, which serves as the basis of theoretical performance guarantees BUCB. Numerical experiments highlight the algorithm's efficiency and demonstrate substantial revenue improvements over single-price benchmarks.

4 - Familiarity-Based Dynamic Pricing with Hierarchical Bayes Estimation of Customer Choices

Binghan Kou, University of Wisconsin, Superior, Superior, WI, United States, Hongmin Li, Yuqi Yang

Repeated consumption affects customers' familiarity with a product and influences their preference for future consumption of the product. This effect varies by individual. For some customers, more familiarity increases their appetite for consumption; for others, the opposite may hold true. More generally, individuals' preferences may be nonmonotonic with respect to familiarity and exhibit various patterns. Accounting for familiarity-based preference creates opportunities for the firm to tailor pricing decisions to each customer's specific preference pattern. A major hurdle, however, is the complexity of deciphering the diverse familiarity-based preference patterns embedded in transaction data and integrating this information into the pricing decision. In this paper, we present a value-passing dynamic pricing strategy that is optimal under a familiarity-based logit choice model. In this context, we develop a versatile finite-state pricing model as an integrated solution that enables not only analytical tractability but also empirical estimation of the familiarity effect. Employing a hierarchical Bayes estimation approach, we demonstrate how this data-driven dynamic pricing strategy improves a firm's bottom line.

5 - Balancing Profitability and Food Waste in Online Grocery: a General Fluid Analysis of Inventory-Adaptive Ranking

Alys Liang, McGill, Montreal, QC, Canada, Huanan Zhang, Sajjad Najafi

Food waste and revenue loss are critical challenges for online grocery retailers, undermining profitability and driving up operational costs. At the retail level, waste is often driven by excess inventory and product expiration. This paper explores how dynamic product ranking can be leveraged to reduce waste by accounting for volatile inventory levels and quality degradation, especially in settings where pricing flexibility and replenishment agility are limited. We demonstrate that adaptive ranking policies can play a vital role in balancing sales performance with sustainability goals.

MC54

Building B Level 2 B216

Innovations in Learning

Invited Session

Revenue Management

Chair: Abdellah Aznag, Columbia University, New York, 10025, United States

Co-Chair: Adam Elmachtoub, Columbia University, NYC, NY, United States

1 - Multi-Group Mean Estimation via Active Learning

Adam Elmachtoub, Columbia University, NYC, NY, United States, Abdellah Aznag, Rachel Cummings

We study a fundamental learning problem over multiple groups with unknown data distributions, where an analyst would like to learn the mean of each group. Moreover, we want to ensure that this data is collected in a relatively fair manner such that the noise of the estimate of each group is reasonable. In particular, we focus on settings where data are collected dynamically, which is important in adaptive experimentation for online platforms or adaptive clinical trials for healthcare. In our model, we employ an active learning framework to sequentially collect samples with bandit feedback, observing a sample in each period from the chosen group. After observing a sample, the analyst updates their estimate of the mean and variance of that group and chooses the next group accordingly. The analyst's objective is to dynamically collect samples to minimize the collective noise of the estimators, measured by the norm of the vector of variances of the mean estimators.

We propose an algorithm, Variance-UCB, that sequentially selects groups according to an upper confidence bound on the variance estimate. We provide a general theoretical framework for providing efficient bounds on learning from any underlying distribution where the variances can be estimated reasonably. This framework yields upper bounds on regret that improve significantly upon all existing bounds, as well as a collection of new results for different objectives and distributions than those previously studied.

2 - Design-Based Confidence Sequences: Stopping Experiments Earlier with Netflix

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

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

applications from Netflix. Our results show that by using our confidence sequence, harmful experiments could be stopped after only observing a handful of units; for instance, an experiment that Netflix ran on its sign-up page on 30,000 potential customers would have been stopped by our method on the first day before 100 observations.

3 - Optimal Offline Policy Learning for Mnl Bandits

Yuxuan Han, New York University, New York, NY, United States, Han Zhong, Miao Lu, Jose Blanchet, Zhengyuan Zhou

Policy learning of the Multinomial Logit (MNL) bandits is a widely studied problem in data-driven assortment optimization, where sellers need to determine the optimal subset of the products to offer based on historical customer choice data. Most existing research on MNL bandits focuses on the online policy learning via repeated interaction with customers, such exploration can be costly in many real-world scenarios. In this talk, we explore the offline learning paradigm by providing a complete characterization of the complexity of offline MNL bandit learning problems. We show that the "optimal item coverage"—where each item in the optimal assortment appears sufficiently often in historical data—is necessary for efficient offline learning by establishing the corresponding information-theoretic lower bounds. For sufficiency, we show that the performance of Pessimistic Rank-Breaking (PRB) algorithm—an algorithm based on pessimism principle for offline decision making—can match the proposed lower bound.

4 - Self-Supervised Post-Training for Better LLM-Agent Decision-Making: a Regret-Minimization Approach

Chanwoo Park, MIT, Cambridge, MA, United States, Ziyang Chen, Asuman Ozdaglar, Kaiqing Zhang

Large language models (LLMs) are increasingly deployed as agents for interactive sequential decision-making tasks. Yet, despite their versatility, LLMs often struggle in dynamic, long-horizon environments such as full-information online learning, multi-armed bandits, contextual bandits, and Markov decision processes—frequently failing to exhibit desirable behaviors like effective exploration or low regret. To address this, we introduce *Self-Supervised Post-Training with Iterative Regret-Minimization*, a self-improving training algorithm that enhances LLM performance in sequential settings. Unlike conventional algorithm distillation, our approach avoids dependence on predefined expert policies or manually crafted language outputs. Instead, it instills a form of adaptive decision-making by jointly optimizing exploration-exploitation dynamics and regret minimization. Empirical results show that *Self-Supervised Post-Training with Iterative Regret-Minimization* consistently improves performance across a spectrum of models—from numerical self-attention agents and lightweight open-source LLMs to the more advanced GPT-4o-mini. Notably, models trained with *Self-Supervised Post-Training with Iterative Regret-Minimization* generalize robustly to out-of-distribution environments, including changes in action/state space size, reward sparsity or shift, and even semantically novel descriptions of real-world sequential tasks. Finally, we provide theoretical insight into why a single-layer attention mechanism—the backbone of many LLMs—can serve as a no-regret learner under our training objective, offering a principled understanding of its decision-theoretic properties.

MC55

Building B Level 3 B314

Data-Driven Decision-Making in Online Marketplaces

Invited Session

Revenue Management

Chair: Chamsi Hssaine, University of Southern California, Marshall School of Business, Los Angeles, CA, United States

1 - The Data-Driven Censored Newsvendor Problem

Sean Sinclair, Northwestern University, Evanston, IL, United States, Chamsi Hssaine

We study a censored variant of the data-driven newsvendor problem, where the decision-maker must select an ordering quantity that minimizes expected overage and underage costs based only on offline censored sales data, rather than historical demand realizations. Our goal is to understand how the degree of historical demand censoring affects the performance of any learning algorithm for this problem. To isolate this impact, we adopt a distributionally robust optimization framework, evaluating policies according to their worst-case regret over an ambiguity set of distributions. This set is defined by the largest historical order quantity (the observable boundary of the dataset), and contains all distributions matching the true demand distribution up to this boundary, while allowing them to be arbitrary afterwards. We demonstrate a spectrum of achievability under demand censoring by deriving a natural necessary and sufficient condition under which vanishing regret is an achievable goal. In regimes in which it is not, we exactly characterize the information loss due to censoring: an insurmountable lower bound on the performance of any policy, even when the decision-maker has access to infinitely many demand samples. We then leverage these sharp characterizations to propose a natural robust algorithm that adapts to the historical level of demand censoring. We derive finite-sample guarantees for this algorithm across all possible censoring regimes and show its near-optimality with matching lower bounds (up to polylogarithmic factors). We moreover demonstrate its robust performance via extensive numerical experiments on both synthetic and real-world datasets.

2 - Learning Fair and Effective Points-Based Rewards Programs

Yichun Hu, Cornell University, Ithaca, NY, United States, Chamsi Hssaine, Ciara Pike-Burke

Points-based rewards programs are a prevalent model to incentivize customer loyalty; in these programs, customers who make repeated purchases from a seller accumulate points, working toward eventual redemption of a free reward. If implemented correctly, they can significantly increase revenue for the seller. However, these programs have recently come under scrutiny due to accusations of unfair practices. Motivated by these concerns, this paper studies the problem of fairly designing points-based rewards programs, focusing on two main obstacles that put fairness at odds with effectiveness. First, customer heterogeneity suggests that different redemption thresholds should be set for different customers to generate high revenue. Second, the relationship between customer behavior and the number of accumulated points is typically unknown, requiring experimentation that may devalue previously earned points.

We first show that an individually fair rewards program using the same threshold for all customers suffers at most a constant factor loss in revenue of $1 + \ln 2$, compared to the optimal personalized strategy. We then tackle the design of temporally fair learning algorithms under demand uncertainty. We propose a “stable” learning algorithm that only changes the redemption threshold $O(\log T)$ times over a horizon of length T , incurring $O(\sqrt{T})$ regret - optimal up to poly-logarithmic factors. We also modify this algorithm to only decrease thresholds, improving fairness at a cost of a constant factor in regret. Finally, extensive numerical experiments reveal the limited value of personalization in average-case settings and demonstrate the strong practical performance of our proposed algorithms.

3 - Satisficing Regret Minimization in Bandits: Constant Rate and Light-Tailed Distribution

Tianyi Ma, Cornell University, Ithaca, NY, United States, Qing Feng, Ruihao Zhu

Motivated by the concept of *satisficing* in decision-making, we consider the problem of satisficing regret minimization in bandit optimization. In this setting, the learner aims at selecting satisficing arms (arms with mean reward exceeding a certain threshold) as frequently as possible. The performance is measured by *satisficing regret*, which is the cumulative deficit of the chosen arm's mean reward compared to the threshold. We propose **SELECT**, a general algorithmic template for **S**atisficing **R**egret **M**inimization via **S**ampLing and **L**ow**E**r **C**onfidence **B**ound **T**esting, that attains constant satisficing regret for a wide variety of bandit problems in the realizable case (*i.e.*, a satisficing arm exists). Specifically, given a class of bandit problems and a corresponding learning oracle with sub-linear (standard) regret upper bound, **SELECT** iteratively makes use of the oracle to identify a potential satisficing arm with low regret. Then, it collects data samples from this arm, and continuously compares the LCB of the identified arm's mean reward against the threshold to determine if it is a satisficing arm. As a complement, **SELECT** also enjoys the same (standard) regret guarantee as the oracle in the non-realizable case. We also study the tail risk control in satisficing regret minimization. We modify **SELECT** and propose **SELECT-LITE** that attains a light-tailed satisficing regret distribution in the realizable case. **SELECT-LITE** also enjoys a constant expected satisficing regret in the realizable case and a sub-linear expected standard regret in the non-realizable case. Finally, we conduct numerical experiments to validate the performance of **SELECT** for several popular bandit settings.

4 - Optimal Incentive Design for Decentralized Dynamic Matching Markets

Pengyu Qian, Boston University, Boston, MA, United States, Chen Chen, Jingwei Zhang

In decentralized dynamic matching markets, agents can benefit from sharing resources, but they will collaborate only if properly incentivized. Motivated by applications such as multi-hospital kidney exchanges and job-hunting markets, this paper proposes novel monetary and non-monetary mechanisms to incentivize collaboration in such decentralized dynamic markets. We study a model in which multiple self-interested agents manage local multi-way dynamic matching problems. Jobs of different types arrive stochastically at each agent, expire after a limited time, and yield rewards when matched. An agent's job backlog and actions are her private information, and each agent aims to maximize her long-run average matching reward.

We design simple mechanisms that incentivize agents to submit all jobs upon arrival, thereby enabling centralized matching. Our first mechanism, the Marginal-Value (MV) mechanism, reimburses agents based on the marginal value of their submitted jobs. This can also be implemented in a non-monetary way by randomly selecting an agent (with a specified probability) to perform a match and collect the associated matching reward. We show that under the MV mechanism, full job submission constitutes an approximate Nash equilibrium—that is, the gain from unilateral deviation vanishes as the number of agents grows. To further eliminate incentives for deviation, we propose a refined mechanism, the Marginal-Value-plus-Credit (MVC) mechanism, and show that when the number of agents exceeds a constant threshold, full job submission constitutes a stronger oblivious equilibrium. Numerical experiments based on kidney exchange data demonstrate that the gains from deviation under our mechanisms are small even in moderately sized markets.

MC56

Building B Level 3 B315

Algorithms and Fairness in Mechanism and Market Design

Invited Session

Auctions and Market Design

Chair: Emily Diana, Tepper, Pittsburgh, PA, United States

1 - Designing Personalized and Equilibrium-Aware Recommendations for School Matching

Erica Chiang, Cornell Tech, New York, NY, United States, Kenny Peng, Nikhil Garg

In New York City, students from underserved communities disproportionately apply to and enroll in high schools with lower metrics such as performance, impact, graduation rate, and selectivity. Existing work suggests this may be due to factors including application complexity, the time cost of forming preferences over many (800+) schools, and unequal access to information—all of which tend to favor more resourced, well-connected families.

Inspired by this setting, we study how to design a recommendation-based informational intervention to help students make more informed decisions about school options. In this talk, I will discuss our approach to designing fair, personalized recommendations that remain effective in equilibrium—the setting where all students receive recommendations at the same time and may change their behavior accordingly.

2 - Improved Algorithms for Fair Matroid Submodular Maximization

Sherry Sarkar, Carnegie Mellon University, Pittsburgh, PA, United States

Submodular maximization subject to matroid constraints is a central problem with many applications in machine learning. As algorithms are increasingly used in decision-making over datapoints with sensitive attributes such as gender or race, it is becoming crucial to enforce fairness to avoid bias and discrimination. Recent work has addressed the challenge of developing efficient approximation algorithms for fair matroid submodular maximization. However, the best algorithms known so far are only guaranteed to satisfy a relaxed version of the fairness constraints that loses a factor 2, i.e., the problem may ask for ℓ elements with a given attribute, but the algorithm is only guaranteed to find $\lfloor \ell/2 \rfloor$. In particular, there is no provable guarantee when $\ell=1$, which corresponds to a key special case of perfect matching constraints.

In this work, we achieve a new trade-off via an algorithm that gets arbitrarily close to full fairness. Namely, for any constant $\epsilon > 0$, we give a constant-factor approximation to fair monotone matroid submodular maximization that in expectation loses only a factor $(1-\epsilon)$ in the lower-bound fairness constraint. Our empirical evaluation on a standard suite of real-world datasets -- including clustering, recommendation, and coverage tasks -- demonstrates the practical effectiveness of our methods.

3 - Minimax Group Fairness in Strategic Classification

Emily Diana, Tepper, Pittsburgh, PA, United States

In strategic classification, agents manipulate their features, at a cost, to receive a positive classification outcome from the learner's classifier. The goal of the learner in such settings is to learn a classifier that is robust to strategic manipulations. We consider learning objectives that have group fairness guarantees in addition to accuracy guarantees. We work with the minimax group fairness notion that asks for minimizing the maximal group error rate across population groups.

We formalize a fairness-aware Stackelberg game between a population of agents consisting of several groups, with each group having its own cost function, and a learner in the agnostic PAC setting in which the learner is working with a hypothesis class H . When the cost functions are separable, we show the existence of an efficient algorithm that finds an approximately optimal deterministic classifier for the learner when the number of groups is small. This algorithm remains efficient, both statistically and computationally, even when H is the set of all classifiers. We then consider cost functions that are not necessarily separable and show the existence of oracle-efficient algorithms that find approximately optimal randomized classifiers for the learner when H has finite strategic VC dimension. These algorithms work under the assumption that the learner is fully transparent: the learner draws a classifier from its distribution (randomized classifier) before the agents respond by manipulating their feature vectors. We highlight the effectiveness of such transparency in developing oracle-efficient algorithms and conclude by verifying the efficacy of our algorithms on real data.

4 - Settling the Competition Complexity of Additive Buyers over Independent Items

Eric Xue, Princeton University, Princeton, NJ, United States

Unlike single-dimensional revenue maximization solved by Myerson (1981), the optimal auction for selling multiple heterogeneous items remains elusive. Over the past two decades, significant progress has been made in designing simple auctions that achieve a constant fraction of optimal revenue. However, a constant fraction may not suffice in some applications, and in others, learning bidder distributions accurately enough is costly.

Bulow and Klemperer (1989) introduced a prior-independent approach, showing that the second-price auction with one extra bidder outperforms the optimal single-item auction without that bidder. Inspired by this, recent work in multi-dimensional settings introduced the notion of *competition complexity*: the number of extra bidders required so that selling items separately with these bidders yields more revenue than the optimal complex mechanism without them.

Previous results determined the competition complexity for n bidders with additive values over $m > n$ independent items as $\Theta(n \log(m/n))$. Our main result settles the open case where $m < n$: we show the competition complexity is $\Theta(\sqrt{nm})$. The $O(\sqrt{nm})$ upper bound was established by Beyhaghi and

Weinberg (2019); we strengthen their lower bound from $\Omega(\ln n)$ to $\Omega(\sqrt{nm})$.

Our proof relies on an explicit construction of a Bayesian IC auction for n bidders with additive values over $m < n$ independent items from the truncated Equal Revenue distribution. This construction achieves revenue exceeding that of selling each item separately to $n + \sqrt{nm}$ bidders, thereby tightly characterizing competition complexity in this regime.

5 - Learning by Climbing: Incentives and Fairness of Upskilling

Guanting Wu, Carnegie Mellon University, Pittsburgh, PA, United States, Peter Zhang, Hai Wang

Problem Definition: Online labor markets (OLMs) are growing rapidly, where platforms recommend matches between clients and workers to ensure project completion and satisfactory *service quality*. Service quality is influenced by worker-project skill compatibility, worker skill levels, and the heterogeneity of their interactions. Due to non-standard match recommendation policies and intrinsic data limitations, quantifying the influence of these determinants on service quality is technically challenging but critical. We propose an empirical approach to understand how workers develop diverse skills by doing different gigs and how such diversity influences service quality, especially when workers focus on a limited set of categories.

Methodology/Results: We develop a partially observable Markov process (POMP) framework instantiated as a multivariate linear-Gaussian model to track latent skill development and its impact on service quality. A two-stage solution approach is proposed to estimate model parameters from a proprietary large-scale dataset with millions of transactions over 12 months. We find that skill development within a category follows a concave growth trajectory, worker-project skill compatibility significantly affects service quality, and project complexity—reflected in contract form, duration, and price—adversely impacts it.

Managerial Implications: Platforms should adopt growth-trajectory-aware and category-aware recommendation systems and consider project decomposition. NLP techniques could analyze dynamic text data to improve service quality. Our estimation framework is also applicable to other matching settings, such as online dating platforms, featuring dynamic and non-standard recommendation policies.

MC57

Building B Level 3 B316

Unlocking the Digital Discourse: From Data to Insights

Invited Session

Social Media Analytics

Chair: Priyanga Gunarathne, University of Pittsburgh, Pittsburgh, PA, United States

1 - The Impact of Premium Licenses on Creator Behavior

Jae Sang Rhee, Georgia Institute of Technology, Atlanta, GA, United States, Mingfeng Lin

Creator economy relies on third-party, free-sharing platforms, which enable creators to reach wide audiences, enhancing monetization opportunities. However, creators remain uncompensated, visibility declines due to oversaturation, and unauthorized use of their work poses risks, especially as AI training datasets frequently source freely accessible content. These issues harm both creators and platforms. Some platforms introduced a premium license, offering subscription-based exclusive content, upfront payments, and enhanced copyright protection. This paper investigates the impact of premium licensing on creator behavior by leveraging a natural experiment. Using data from Unsplash and Pexels, we find that premium licenses on free-sharing platforms reduce the volume of freely available content by 13.2%, even among creators without license access who are not directly affected. We identify two mechanisms driving this decline. First, reduced multi-homing occurs as creators deactivate accounts and move away from the licensed platform. Second, creators improve free content quality to compete with premium offerings. These trades-offs reveal unintended effects on both content volume and quality, affecting creators and platforms alike. Consequently, effective monetization policies must balance platform revenue with a healthy content ecosystem and creators' branding goals. Premium licensing signals a broader shift from open to closed access, reshaping how content is produced and consumed. As AI heightens the need to safeguard creative work, understanding these dynamics becomes increasingly urgent, calling for further research into premium licensing and its role in the evolving creator economy.

2 - How Prospective Accounting Bias Impacts Streaming Media Consumption

Yifei Wang, University of New Hampshire, Durham, NH, United States, Maya Mudambi

In the streaming era, consumers can subscribe to platforms like Netflix, which offer access to a library of titles for a subscription fee, or rent/purchase individual titles. According to prospective accounting theory (Prelec & Loewenstein, 1998), payment method and timing shape consumer behavior. This study examines how subscription models influence consumption on a platform offering both payment structures, contributing to research on mental accounting, streaming, and platform strategy. Using a staggered difference-in-differences approach, we exploit the timing of films shifting from rental-only to subscription. Subscription access increases overall consumption and exploratory viewing but lowers average ratings. Content led by women and people of color sees a greater viewership boost, suggesting stronger engagement with marginalized voices. Robustness checks support prospective accounting bias as the key mechanism, ruling out alternative explanations. Future work will further quantify exploratory viewing and examine how pricing models interact with historical media preferences across multiple dimensions.

3 - From Disruption to Dialogue: Professional Discourse on Social Media on Emerging Technologies

Yupei Hu, University of Pittsburgh, Pittsburgh, PA, United States, Priyanga Gunarathne, Xiaomeng Chen, Narayan Ramasubbu

This study explores the evolution of professional discourse around a disruptive technology within a physician community on X (formerly Twitter), focusing on its emergence and progression toward an AI-empowered medical specialty. Drawing on a comprehensive dataset of all professional interactions among physicians in this community over a six-year period, we adopt a mixed-methods approach—combining social network analysis, topic modeling, and temporal discourse analysis—to trace how this technology entered professional conversations, gained momentum, and ultimately intersected with narratives around artificial intelligence. By mapping the dynamics of influence, adoption, and thematic shifts, the study offers novel insights into how disruptive technologies diffuse and evolve within professional digital communities. The findings contribute to the literature on digital transformation in healthcare, technology adoption among professionals, and the role of social media in shaping innovation trajectories.

4 - Strategic Visuals: Metaphorical Design and Viewer Emotion in Vi Marketing

Ying Wang, Northern Illinois University, Naperville, IL, United States, Jaeki Song

As virtual influencers (VIs) become increasingly prominent in digital marketing, understanding how their visual design influences audience emotions is essential. Drawing on Conceptual Metaphor Theory and Visual Rhetorical Theory, this study investigates how different types of visual metaphors affect emotional engagement in VI-based video campaigns. We analyzed 100 YouTube videos featuring virtual influencers using computer vision (YOLO) to classify visual metaphors and natural language processing to assess viewer sentiment and influencer-product alignment. This study contributes to marketing and information systems fields by extending metaphor research into dynamic video contexts and highlights the critical role of symbolic coherence and design in VI marketing. Practically, the study offers actionable insights for brands and content creators on how to craft emotionally compelling virtual influencer content through effective metaphorical storytelling and strategic alignment.

MC58

Building B Level 3 B317

Manufacturing Process Optimization

Contributed Session

Contributed

Chair: JinGoo Jeong, Kyung Hee University

1 - Optimization Models for the Sand-Casting Processes in Foundries

Hajar Zegrati, University of Brescia, Brescia, Italy, Luca Bertazzi

We study a complex real-life industrial scheduling problem concerning the melting process in a foundry industry using sand casting. For this problem, we formulate a mixed integer linear programming formulation able to overcome a previous non-linear model. The main challenge is the complexity in handling the residual liquids that are reused in subsequent castings if the same type of alloy is being produced. The setup costs are represented by the change of alloy that requires the emptying of the furnace from the previously processed alloy and therefore depend on the sequence of production. Moreover, we introduce goal programming to find the best balance between different objectives, in terms of total cost, energy consumption, and pollution, considering that foundries are responsible for one third of global energy consumption and a quarter of greenhouse gas emissions. Extensive computational results show that the new model is able to find better solutions than the previous one in short computational time and better balanced solutions in terms of total cost, energy consumption, and pollution.

2 - An Adp Algorithm for Dynamic Steel-Making Scheduling Problem Considering Ladle Assignment

Yuxuan Zhang, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, China, Shenyang, China, People's Republic of, Ying Meng, Lixin Tang

In steel-making workshops, heats are scheduled according to their production routes. The feasibility of the schedule is greatly affected by the uncertain availability of ladles, the containers of molten steel during production. In the worst cases, the production procedure would even be interrupted. In this paper, we investigate a dynamic steel-making scheduling problem considering the feasible assignment-timetabling of ladles under uncertain production durations. To illustrate the problem, we develop an MIP formulation based on perfect information. While the existing research literature has been focusing on finding a dynamic strategy that minimizes the expected value of the objectives enacted, our problem emphasizes the performance under the worst-case scenario in this stability-focusing production environment. Specifically, the problem is formulated as a Markov decision process that applies a robust optimization paradigm at each decision epoch. To address this problem, we develop an effective approximate dynamic programming (ADP) algorithm, which relies on look-ahead simulations based on Monte Carlo sampling and a decision rule-based ladle assignment strategy. Computational results on the random instances show that our ADP algorithm can obtain competitive solutions with the greedy algorithms in a reasonable computational time and prove the robustness against uncertainty.

3 - Exact Algorithm for Plate Design Problem with Uncertain Trim Loss in Steel Enterprises

Sen Li, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, China, Shenyang, China, People's Republic of, Ying Meng, Lixin Tang

In the steel plate production process, slabs are rolled into wide and thin raw plates, which are then cut into smaller plates to meet customer demands. Effective plate design is critical for improving material utilization. However, fluctuations in equipment performance usually introduce uncertainty in trim loss, leading to deviations from the design plan and resulting in material waste. We study a plate design problem under uncertain trim loss, which is modeled as a two-dimensional, two-stage guillotine cutting stock problem with knapsack-type uncertainty. A budget uncertainty set is adopted to describe the uncertain trim loss. Then, we formulate the problem as a set covering model, which is further reformulated into a deterministic equivalent. An adapted branch-and-price algorithm is developed for the obtained deterministic cutting stock problem. Furthermore, we propose a tailored solution approach for the two-stage subproblem and design an exact labeling algorithm to solve it optimally. Computational experiments demonstrate that our method significantly improves material utilization and enhances robustness compared to traditional approaches.

4 - A Two-Stage DRO Approach for Production Planning with Correlated Machine Breakdown

Qingyang Wang, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, China, Shenyang, China, People's Republic of, Ying Meng

We develop a two-stage distributionally robust optimization (DRO) framework for the production planning problem under uncertain machine breakdowns (PPUMB). In the first stage, material allocations are determined; in the second, production plans are optimized in response to potential machine breakdown scenarios. Instead of assuming independent machine breakdowns, we consider correlated breakdowns across machines and model machine breakdowns using joint probability distributions, which better reflects real-world production environments. We prove that the second-stage expected objective function is supermodular, and leverage this structural property to derive the worst-case distribution and reformulate the DRO model into a deterministic equivalent. This reformulation introduces substantial computational complexity due to the increase in variables and constraints. To address this, we design a tailored multicut Benders decomposition (MBD) algorithm. Additionally, we incorporate supervised learning into the solution process by training a classifier to identify and retain only high-quality optimality cuts, thereby improving computational efficiency. Experiments based on real production data demonstrate that the proposed approach significantly reduces solution time and yields robust, high-quality production plans under uncertainty.

5 - Integrated Order Acceptance, Production Scheduling, and Distribution In Additive Manufacturing with Energy Consumption and Material Recycling Considerations

Fatemeh Pourdehghan, University of Missouri, Columbia, MO, United States, Omid Shahvari, Sharan Srinivas

This study addresses a problem of integrated order acceptance, production scheduling, and delivery routing for additive manufacturing (AM) systems using selective laser melting (SLM). AM shifts production from "make-to-stock" toward customized and make-to-order items, printed only after an order is confirmed. Integrating production and delivery in 3D printing is essential, as ignoring delivery constraints causes production scheduling to focus solely on production objectives, like grouping orders by material. This results in high delivery costs if customers are far apart. A key advantage of SLM is its ability to reuse remaining material from previous batches, enhancing material efficiency and reducing waste. By considering material-specific refresh rates, we aim to improve the overall sustainability and cost-effectiveness of the process. The production stage involves batch processing of parts on unrelated parallel AM machines with material eligibility constraints. Once production is complete, parts are regrouped into delivery batches- independent of the initial batches- and delivered to customers using heterogeneous vehicles. Each vehicle performs one trip, and routing follows vehicle routing problem (VRP) framework. The objective is to maximize total profit, defined as total revenue minus relevant costs, including material waste, setup, tardiness, energy consumption, and transportation. To the best of our knowledge, this is the first study to jointly optimize order acceptance, AM-based production scheduling, and VRP, while incorporating energy consumption and material recycling. We propose a mixed-integer linear programming (MILP) model for small-sized instances and an Adaptive Large Neighborhood Search (ALNS) for large-sized problems. Computational experiments highlight the effectiveness of the proposed algorithm.

6 - Balancing Risk and Repairs: A Consumer-Driven Model for Electronics Repair Support Duration

Meltem Denizel, Iowa State University, Ames, IA, United States, Haley Paluzzi

Designing electronics with repairability in mind can extend product lifespans, which improves resource efficiency and promotes sustainability. However, repair rates for used electronics in the United States remain low, with most consumers choosing to replace damaged items rather than repairing them. The purpose of this study is to explore how consumer perceptions of risk influence their willingness to engage in product repairs. By understanding how willingness to pay for repairs changes over time, we can develop a model to predict the optimal duration for offering repair services. Therefore, this study seeks to address two research questions: (1) What is the nature of the relationship between the expected product lifespan and willingness to pay for product repairs over time? and (2) How do decreasing rates in willingness to repair impact the ideal length of time that repair services should be offered?

7 - Power Simulation for Milling Based on a Generalized Model Incorporating Tool Wear and Material Properties

JinGoo Jeong, Kyung Hee University, Yongin-si, Korea, Republic of, Ga Hyun Lee, Byunghun Song, Hyun Woo Jeon

Enhancing energy efficiency in manufacturing is critical for both addressing climate change and optimizing operational performance. Among various machining processes, milling is particularly energy-intensive, making accurate power prediction essential for effective energy management. Traditional power models based on material removal rate (MRR) often disregard tool wear, a significant driver of power variation over time. While several prior studies have attempted to incorporate tool wear, their applicability is frequently limited to specific materials or experimental setups. To overcome these limitations, this study proposes a dynamic power model that generalizes across diverse manufacturing conditions by integrating tool wear and workpiece material hardness as core variables. We implement this model in a discrete event simulation (DES) framework and evaluate power consumption time series across three distinct workpiece materials—aluminum alloy, stainless steel, and titanium alloy. The simulation results indicate that titanium alloy, the hardest material, leads to the highest average power and energy consumption, while aluminum alloy, the softest, shows the highest peak power due to prolonged tool wear effects. In addition, the results are analyzed using multiple comparison criteria such as replacement frequency and power fluctuations. These findings demonstrate the complex interaction between tool wear and material hardness in shaping power demand. The proposed model facilitates more realistic and scalable energy analyses, contributing to informed decision-making for sustainable manufacturing.

MC59

Building B Level 3 B318

Healthcare Operations in Pandemics

Contributed Session

Contributed

Chair: Samira Alsadat Salari, University of Arkansas, 970 N Oakland Ave Apt 10, Fayetteville, AR, 72701-1925, United States

1 - Catch Me if You Can: An Optimal Pest Survey Strategy to Delimit Biological Invasions

Spencer Hill, Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada, Jue Wang, Denys Yemshanov

We consider the problem of optimal allocation of limited resources to delimit the frontier of the spread of an invasive species population. Given a fixed sampling budget and an invasive species detection rate and linear population spread rate, we formulated the delimiting survey problem as a partially observable Markov decision process (POMDP) where the objective is to minimize the cumulative gap between the unobservable (true) infestation frontier and the closest location with a detected infestation. We proved that the resulting POMDP problem can be equivalently formulated and solved as a deterministic optimal control problem. In the simplest case of two sampling locations, we used Pontryagin's maximum principle to describe the optimal sampling strategy. The optimal control is heavily influenced by the survey budget level; for any budget level greater than twice the ratio between the spread and detection rate, it is optimal to sample only at the closest site to the previously detected infestation. In a general case of N survey sites, we applied a discrete-time numerical approximation to compute a near-optimal sampling strategy. We evaluated the surveillance strategies for four species with distinct combinations of detection and spread rates. Our work provides pest management practitioners with practical insights for developing heuristic sampling strategies to maximize the capacity to delimit the spreading pest invasion frontier.

2 - Multi-Stakeholder Public-Private Partnerships (MSPs) in Health Emergencies

Ioanna Falagara, Vienna University of Economics and Business, Vienna, Austria, Evelyne Vanpoucke, Adam Abdelsalam

Multi-Stakeholder Partnerships (MSPs) play a vital role in responding to global health emergencies. By leveraging the complementary strengths of the public and private sectors and collective actions, MSPs help ensure the timely and equitable distribution of essential medical resources during health emergencies.

This research investigates how stakeholder salience shapes the structure and function of MSPs formed in response to global health crises. Applying the three dimensions of stakeholder salience—power, legitimacy, and stake—the research evaluates how various actors influence the formation and operational dynamics of MSPs. The findings highlight the diverse contributions of stakeholders and reveal how disparities in salience impact decision-making processes, governance structures, and resource allocation within MSPs.

3 - Assessing the Assumptions of Vehicle Dispatch Timing in a Disaster-Relief Supply Chain Network

Robert Cook, Northeastern Illinois University, Chicago, IL, United States

This study assesses the impact of one key assumption from the seminal work in the Disaster Relief Logistics literature regarding the timing of dispatch decisions between a Staging Area and the Point of Distribution in a Last Mile Distribution context. We compare versions of the model where beneficiaries wait until the end of a decision period before leaving the Point of Distribution and where they vacate immediately after receiving as much inventory is available. Similar to previous studies, continuously dispatching the vehicle (CD) is shown to be an optimal or nearly-optimal policy in the majority of conditions tested, and even narrowed the gap to the optimal policy previously shown in literature. Therefore, CD is the best available decision policy for practical use.

4 - Dynamic Outbreak Countermeasure Logistics

Mohsen Mehrabiyan, West Virginia University, Morgantown, WV, United States, Zeyu Liu

The effective deployment of medical countermeasures (MCMs) is critical to strengthening national readiness against epidemic outbreaks. However, uncertain demand, supply shortage, and unequal distribution of healthcare resources have complicated MCM deployment planning for practitioners. In this study, we develop a multi-stage stochastic programming (MSP) model to optimize facility location, resource allocation, and dynamic MCM deployment decisions under uncertainty. We further improve the MSP model by embedding vehicle routing constraints and equitable resource distribution constraints. Due to the complexity of solving large-scale problems, we employ Nested Benders Decomposition (NBD) and design an improved Progressive Hedging (PH) algorithm. Artificially generated outbreak scenarios are used as benchmarks to test algorithm performances. The model is then calibrated with real-world COVID-19 transmission and capacity data from the state of West Virginia to evaluate the practical impact of vehicle routing and equitable distribution decisions. The case study highlights clear trade-offs between delivery speed, cost, and fairness, offering practical guidance to public health practitioners.

5 - Designing a Sustainable Vaccine Supply Chain by Considering Demand Substitution and Value-Added Function During a Pandemic Outbreak

Samira Alsadat Salari, University of Arkansas, Fayetteville, AR, United States

Pandemics like COVID-19 impose serious social, economic, and health burdens, prompting governments to enhance national resilience through effective vaccination strategies. Some countries rely on both domestically produced vaccines and international aid, necessitating the development of a sustainable and responsive vaccine supply chain. This study incorporates vaccine-specific efficacy and shelf life, based on established literature, to address the perishability and performance of different brands. Recognizing that vaccine shortages are common during pandemics due to procurement, distribution, and storage constraints, we examine demand substitution based on relative efficacy to mitigate unmet demand and reduce infection rates. The model also integrates both permanent vaccination sites (e.g., hospitals and clinics) and mobile centers to improve accessibility and coverage across diverse geographic and demographic populations. A multi-objective mathematical model is formulated and solved using the augmented ϵ -constraint method. A real-world case study demonstrates the model's applicability. Results and sensitivity analysis reveal that combining fixed and mobile vaccination centers enhances service delivery, while demand substitution offsets approximately 60% of shortages, accelerating immunization efforts.

MC60

Building B Level 3 B319

Generative AI Strategy and Implementation

Contributed Session

Contributed

Chair: Timothy Anderson, Portland State University, Portland, OR, United States

1 - Accuracy and/OR Explainability? the Deployment and Adoption of Generative Ai with Hallucinations**Xu Zhang, School of Management, Fudan University, Shanghai, China, People's Republic of, Siyi Yang, Yifan Dou, Cheng Zhang**

Generative AI tools promise significant productivity gains but are hindered by a persistent challenge: **hallucinations**—outputs that include false or misleading information presented as facts. This paper examines how organizations can strategically deploy AI tools when such hallucinations exist, particularly focusing on two key levers: improving the **model accuracy** (reducing hallucination frequency) and enhancing the **output explainability** (lowering verification costs). We develop a game-theoretic model in which a principal provides generative AI to an agent tasked with completing a quality-sensitive assignment. The agent decides whether to adopt the AI, whether to verify its output, and whether to invest further effort in improving its quality. We show that accuracy and explainability share a dynamic, non-monotonic relationship: when accuracy is low, explainability serves as a substitute by reducing verification cost and encouraging responsible AI use. When accuracy is high, the two act as complements, jointly necessary to avoid free-riding behavior. These findings challenge the conventional wisdom that explainability and accuracy can be treated as simple substitutes. We further consider some forms of explainability provide supportive justifications for AI-generated content—regardless of whether it is correct—which may mislead users' judgment. We examine how users with varying discernment capabilities respond to Gen-AI outputs accompanied by the explanations. Our results inform how firms should balance technical improvements with user-interface design to achieve high-quality outcomes. By modeling both sides of AI deployment—technical design and user adoption—our research offers managerial supports for deploying trustworthy AI and contribute to the economic understanding of technology adoption.

2 - Unlocking the Value of Custom AI: The Role of Proprietary Resources**Xiaoyang Zhen, IE Business School, Madrid, Spain, Fabrizio Salvador**

Large language models are spawning a new generation of custom AI agents that combine generic model capabilities with task specific data, tools, and workflows. Yet we know little about whether the proprietary resources — domain knowledge bases, private plugins, and structured workflows—actually make these agents more valuable. We address this gap by analyzing over 400,000 agents created on an AI platform, one of the world's largest no code agent platforms. Leveraging platform telemetry with detailed agent-configuration records, we estimate the causal effect of proprietary resource adoption on agent success using an instrument variable strategy and category fixed effects models. This study aims to contribute to the platform ecosystem literature by highlighting proprietary resources as critical yet underexplored factors in decentralized AI development. The insights offer practical implications for developers and platform managers in designing resource integration strategies to optimize the value generation of AI agents.

3 - Decoding Individuals' Resistance Toward GAI Adoption in the Workplace**Neda Khosh Kheslat, University of North Texas, Denton, TX, United States**

As one of the most revolutionary technologies, generative AI adoption in an organizational setting is essential for improving performance and raising work productivity. However, previous studies in the change management field have consistently shown that user resistance plays a significant role in adopting new information technology. This study aims to determine factors influencing employee resistance to adopting GAI technologies in the workplace through an integrative framework inspired by innovation resistance theory (IRT) and the status quo bias perspective. This study also takes a unique and nuanced approach regarding the function of GAI models capable of autonomy, decision-making, and learning.

4 - Perceptions and Usage of Gen AI Chatbots - A Meta-analytic Systematic Review**Suchitra Veera, Snayu, Fort Worth, TX, United States, Pranav Veera**

Chat GPT and similar Generative Artificial Intelligence (Gen AI) technologies (Llama, Gemini, etc.) that are based on Large Language Models (LLMs) have been gaining rapid adoption in some areas of business recently. The use of chatbots is one of the highlights of the implementation of these Gen AI models for serving the needs of customers for companies in the roles of call center agents, marketing content developers, as well as for educational tutoring. This presentation will discuss the findings from a systematic review study of previous empirical research to develop a better understanding of how Gen AI usage is gaining adoption in the form of a variety of chatbot-based products and services. Using a meta-analytic approach this research will provide insights into the perceptions and usage of Gen AI technology-based products such as chatbots. The meta-analysis is based on data gathered from prior empirical studies conducted by researchers over the past five years. The results from the findings of the meta-analysis on the perceptions and usage of chatbots provide details of the variability of effect sizes across studies, estimate of the population effect size with associated confidence interval, moderator variables that can explain the variability, and publication bias in the studies. Such an analysis provides an understanding of the areas of growth and demand in the use of chatbots and use cases where adoption is lagging. These insights can help to steer the development of such chatbot-based products that are more effective and serving the needs of customers better.

5 - Technology Forecasting of AI LLMs**Timothy Anderson, Portland State University, Portland, OR, United States**

Presentation will provide initial results of using TFDEA, Technology Forecasting Using Data Envelopment Analysis of generative AI LLM platforms.

MC61

Building B Level 4 B401

Small-Sample Machine Learning and its Applications in Smart Manufacturing

Invited Session

Quality, Statistics and Reliability

Chair: Weizhi Lin, San Jose State University, San Jose, CA, United States

Co-Chair: Qiang Huang, University of Southern California, Los Angeles, CA, United States

1 - Quantum Bayesian Optimization for Quality Improvement in Fuselage Assembly**Jiayu Liu, Rensselaer Polytechnic Institute, Green Island, NY, United States, Yinan Wang**

Recent efforts in smart manufacturing enhanced aerospace fuselage assembly processes, particularly by innovating shape adjustment techniques to minimize dimensional gaps between assembled sections. Existing approaches, including reinforcement learning or Bayesian optimization with surrogate models, have shown promising results but face limitations such as sensitivity to measurement noise and being extremely data-demanding (low sample efficiency). The development of quantum computing shed light on revolutionarily resolving the aforementioned challenges. We propose a Quantum Bayesian Optimization (QBO) framework for precise shape control during assembly. Specifically, this approach utilizes a quantum oracle based on finite element analysis (FEA)-based models or surrogate models, to acquire more accurate observations with fewer queries for a certain input. QBO employs an Upper Confidence Bound (UCB) acquisition function to strategically select input values that are most likely to minimize the gap. It has been theoretically proved to require much fewer samples while maintaining comparable optimization results. In the case study, QBO is applied to optimize force-controlled actuators for adjusting the shape of fuselage sections and reducing the gap to the adjoining section. Experimental results demonstrate that QBO achieves apparently lower cumulative regret compared to classical methods, particularly under noisy measurement conditions, with significantly fewer samples queried from the simulation.

2 - Adaptive Generative Modeling with Continual Learning for Data-Scarce Manufacturing**Mengfei Chen, The State University of New York at New Paltz, New Paltz, NY, United States, Vidita Gawade**

Manufacturing processes continually evolve to accommodate new products, tool variations, and shifting operating conditions, resulting in mixed, heterogeneous, and often anomalous data. This variability makes it difficult for static quality prediction models to generalize across tasks. A key challenge is data scarcity, caused by the infrequency of defects and the constant introduction of new product types, leading to limited historical data and highly imbalanced datasets, both of which degrade model

performance. While data augmentation is often used to address imbalanced data, many methods produce physically unrealistic synthetic data, undermining predictive accuracy despite advanced model architectures. To address these challenges, this study proposes an adaptive generative modeling framework integrated with continual learning. The approach synthesizes physics-constrained, physically valid in-process sensor data to balance datasets while employing continual learning in the generative model to adapt to evolving manufacturing scenarios. The method is validated through two case studies: porosity prediction in laser metal deposition and quality prediction through thermal images in the hot stamping process.

3 - Physics-Informed Cross-Modality Transfer Learning in Diffusion Models for Data Augmentation in Metal-Based Additive Manufacturing

Wenmeng Tian, Mississippi State University, Mississippi State, MS, United States, Hugo Rodriguez

Anomaly detection and in-situ monitoring are essential in additive manufacturing (AM) quality control. Artificial intelligence (AI) approaches can be used to leverage AM process data for quality improvements. However, limited data availability is one of the challenges in the applications of AI techniques in AM. Generative models (GMs) can be leveraged for data augmentation, enabling model performance improvement. Nevertheless, these models still require a sufficiently large dataset to generate high-quality synthetic samples. In AM, data collection can be very costly and time consuming. Therefore, transfer learning (TL) can be used to leverage publicly available datasets as source datasets, to transfer the knowledge obtained from the source domain to the target domain with a small-scale dataset. However, the high diversity of sensors for data collection in both source and target domains makes the transfer of knowledge across different sensor modalities more challenging. Therefore, a cross-modality transfer learning approach is proposed by leveraging a publicly available dataset of AM melt-pool optical images to improve the generative capabilities of a Denoising Diffusion Implicit Model (DDIM) by implementing model-based TL for data augmentation of AM thermal history. The real-world case study shows that the proposed approach enhances the performance of thermal image generation, especially when there is limited target data available.

4 - Domain-Informed Spatial Function Prediction on Surface Manifolds for Capturing Manifold-to-Manifold Variability

Weizhi Lin, University of Southern California, Los Angeles, CA, United States, Qiang Huang

Models for predicting spatial functions typically capture the correlations between covariates and scalar fields defined on a fixed surface manifold. However, the embeddings of surface manifolds may vary in practice, including both their intrinsic geometries and extrinsic positions in three-dimensional space. This manifold-to-manifold variability introduces complex spatial patterns and challenges accurate modeling of covariate effects with limited data. To address this, we propose decomposing spatial functions to characterize the variability introduced by intrinsic geometry and extrinsic position. The variability driven by intrinsic geometry is captured by the number and locations of critical points, which correspond to local high-order spatial patterns and remain invariant to extrinsic positioning. Meanwhile, the variability induced by extrinsic position is captured by local high-order spatial patterns whose locations are unrelated to the surface's internal structure. The shapes and locations of these high-order patterns are automatically identified and estimated using domain-informed adaptive Gaussian kernel bases. To enable prediction of complex spatial functions on a new manifold, we establish separate models for the intrinsic-driven, extrinsic-driven high-order patterns and smooth low-order patterns. For each component, a greedy best subset selection is performed to identify the most relevant covariates, which are then modeled with a generalized functional additive mixed model. We demonstrate the effectiveness of the proposed methods through a real-world case study on modeling surface deviations of 3D-printed objects. With limited samples, the method accurately predicts complex surface deviation patterns for parts printed at different locations on the print bed.

5 - Identification of Latent Invariant Surface Quality Patterns via Spatial Stochastic Process Deformation

Minghao Gu, University of Southern California, Los Angeles, CA, United States, Qiang Huang

Surface quality characterization is important for ensuring product functionality and guiding production. Although a wealth of literature has been developed for mass production, unique challenges arise in additive manufacturing (AM) because of the one-off fabrication of a wide range of complex geometries. Surface quality patterns can change with built geometries and covariates such as shape size and built location. Existing statistical characterization methods are therefore not adaptive to AM due to frequent design changes and insufficient samples for each design. Since a product can be represented by a finite type of surface patches, this initial study investigates the surface quality of products with spherical patches. We propose to identify latent and invariant patterns to characterize surface quality by separating the effect of covariates. The surface quality of different domes is first described by nonstationary spatial Gaussian processes. These heterogeneous spatial stochastic processes are then mapped to one underlying base process through a spatial stochastic process deformation method. With covariate effects fully captured by the mapping function, the underlying base process characterizes the latent invariant surface quality patterns of spherical patches and the working status of the machine. Actual printed samples from an AM process are utilized for methodology demonstration. The study brings the prospect of evaluating and comparing product surface quality and process condition in one-off AM.

MC62

Building B Level 4 B402

Reinforcement Learning for Process Control and Maintenance

Invited Session

Quality, Statistics and Reliability

Chair: Ying Fu, University of Wisconsin-Madison, Madison, WI, United States

Co-Chair: Kaibo Liu, UW-Madison, Madison, WI, United States

1 - Sequential Conformal Prediction for TIME Series with Uncertainty Quantification

Hanyang Jiang, Georgia Institute of Technology, Atlanta, GA, United States

Conformal prediction (CP) has been a popular method for uncertainty quantification because it is distribution-free, model-agnostic, and theoretically sound. For forecasting problems in supervised learning, most CP methods focus on building prediction intervals for univariate responses. In this work, we develop a sequential CP method called MultiSPCI that builds prediction regions for a multivariate response, especially in the context of multivariate time series, which are not exchangeable. Theoretically, we estimate finite-sample high-probability bounds on the conditional coverage gap. Empirically, we demonstrate that MultiSPCI maintains valid coverage on a wide range of multivariate time series while producing smaller prediction regions than CP and non-CP baselines.

2 - Deep Belief Updater Model for Coupled Higher-Order Hidden Semi-Markov Chains

Vipul Bansal, University of Wisconsin Madison, Madison, WI, United States, Shiyu Zhou

We propose a Deep Belief Updater (DBU) Model to address the challenge of modeling complex temporal dependencies in Coupled Higher-Order Hidden Semi-Markov Chains. These models extend Higher-Order Hidden Semi-Markov Chains by incorporating temporal dependencies among hidden states across multiple interdependent chains. Replacing underlying temporal dynamics with neural networks is computationally intensive, as it requires updating the state history at each time instance for each chain, based on the history of all the chains. To overcome this, we find that forward-state probabilities, the probabilities of being in a given state at a time instance based on the history of observations, can be expressed as a function of the Stage I belief states derived from individual observations. Stage I belief state is the likelihood of the system being in a certain state based solely on the current observation, without considering the influence of previous or future observations. The DBU model directly estimates these forward-state probabilities using Stage I belief. These probabilities are used to construct a likelihood function to train the model. By enabling the direct estimation of forward-state probabilities, the DBU model effectively captures the intricate transition dynamics of systems with interdependent chains.

3 - Degradation Control and Maintenance via Reinforcement Learning

Ying Fu, University of Wisconsin-Madison, Madison, WI, United States

Many industrial systems, such as manufacturing and energy facilities, consist of a fleet of units with redundancy that work together to meet production goals. As these units degrade over time, they require maintenance to keep the system operational. However, maintenance takes time and is often constrained by limited resources. During maintenance, the remaining units must take on additional workload to maintain overall productivity. Since degradation typically depends on workload, there is an opportunity to reduce maintenance costs and prevent failures by strategically distributing the operational workload across the fleet. For example, when maintenance capacity is fully utilized, assigning less workload to units nearing failure can slow their degradation and extend their lifespan. This study proposes a reinforcement

learning (RL) approach to jointly optimize workload allocation and maintenance scheduling. A case study using Raspberry Pi devices demonstrates the effectiveness of the proposed method.

4 - Variance-Aware Risk-Sensitive Reinforcement Learning

Saunak Kumar Panda, University of Houston, Houston, TX, United States

We propose a variance-penalized reinforcement learning (RL) framework that incorporates empirical uncertainty estimates to improve the reliability and stability of learned policies. Unlike prior methods that rely on auxiliary critics to approximate second moments of returns, our approach directly penalizes variance using non-parametric, statistically principled techniques such as bootstrapping and random scaling. These methods allow for efficient online variance estimation without additional value networks, enabling lightweight and scalable risk-sensitive learning. We instantiate this framework in both tabular Q-learning and a policy-gradient actor-critic algorithm, introducing empirical variance penalties into standard updates without compromising convergence. Empirically, our approach consistently achieves more stable policy behavior in stochastic domains, such as the noisy puddle world environment. Across discrete and continuous settings, the empirical variants match or outperform prior variance-penalized baselines in terms of return stability while retaining competitive expected performance. Our results highlight the potential of embedding non-parametric uncertainty estimation into the risk-sensitive RL paradigm without increasing model complexity.

MC63

Building B Level 4 B403

Statistics and Process Control

Contributed Session

Contributed

Chair: George Mytalis, Athens University Economics Business, Hlioupoli, Greece

1 - Repair-Aware Redundancy Allocation: A Hybrid Simulation and Metaheuristic Approach

Berna Dengiz, Baskent University, Ankara, Turkey, Merve Uzun, Orhan Dengiz

The Redundancy Allocation Problem (RAP) is important for designing reliable systems. In this study, we aim to find the best design for a series-parallel system by considering more realistic conditions. Unlike traditional methods that assume failed components cannot be fixed, we consider repairable components with increasing failure and repair rates that change over time. This makes the system more efficient and reduces unnecessary cost and weight.

We use a Discrete Event Simulation (DES) model together with Monte Carlo (MC) Simulation to estimate the reliability of the system. To solve the RAP, we apply two metaheuristic algorithms: a steady state Genetic Algorithm (ssGA) and a Memetic Algorithm (MA). The MA improves the ssGA by adding a smart local search method that uses problem knowledge.

Our results show that when repairable components are included in the model, the system can be designed with fewer spare parts, less weight, and lower cost while still maintaining high reliability. The MA also finds better solutions than the ssGA in the same amount of computation time.

2 - Nonintrusive Device-Level Pattern Recognition and Power Consumption Prediction in Buildings Through Sequential Variational Autoencoders

Salih Salihoglu, University of Miami, Coral Gables, FL, United States, Ramin Moghaddass

The accurate prediction and disaggregation of energy consumption at the device level in residential and commercial buildings remain critical for effective energy management and demand response. This study introduces a novel framework for nonintrusive load monitoring (NILM) by leveraging a deep generative modeling approach using sequential variational autoencoders (VAEs). Our method reframes appliance-level power disaggregation and forecasting as a dynamic state-space modeling problem, addressing key challenges such as overlapping device usage patterns, temporal dependencies, and data sparsity. The proposed model captures both the stochastic nature and temporal evolution of device usage through an autoregressive latent variable architecture. We incorporate long-term dependencies and inter-device interactions into the latent space without requiring intrusive sensing or labeled training data. Evaluations using generated and publicly available datasets demonstrate the model's performance in power consumption prediction compared to conventional NILM techniques. This work represents a significant step toward scalable and interpretable solutions for smart grid applications and energy-efficient building management.

3 - A Hybrid Isolation Forest and LSTM Variational Autoencoder Approach for Phase I Statistical Process Control

Waldyn Martinez, Miami University, Oxford, OH, United States

Phase I statistical process control (SPC) primarily involves analyzing historical data to detect outliers, assess process stability, and establish robust control limits for future monitoring. Traditional SPC methods often assume independence and stationarity of observations—assumptions frequently violated in complex industrial processes with autocorrelated or multivariate data. To address these limitations, we propose a novel hybrid methodology integrating Isolation Forests (IF) with a Long Short-Term Memory Variational Autoencoder (LSTM-VAE). The LSTM-VAE leverages temporal dependencies inherent in sequential data, capturing complex nonlinear dynamics and refining the distributions of the latent space for potentially more robust anomaly detection through reconstruction error analysis. Subsequently, the Isolation Forest provides the anomaly detection capability by effectively isolating anomalies without strong assumptions about data distribution. This hybrid approach significantly enhances detection accuracy and robustness, particularly in processes exhibiting temporal autocorrelation or intricate multivariate relationships. We demonstrate the effectiveness and practical applicability of our proposed method using simulated and real-world industrial datasets, highlighting its superior performance in accurately identifying anomalies and setting appropriate Phase I control limits.

4 - Productivity Change Estimation Under Data Uncertainty: a Forest-Based Probabilistic Approach

Yu Zhao, Tokyo University of Science, Shinjuku-ku, Japan

This study introduces a novel method for estimating productivity change that explicitly accounts for uncertainty in input-output data through the Malmquist Productivity Index framework. By applying a forest-based clustering and sampling procedure, we capture the underlying probabilistic structure of observed decision-making units (DMUs). Clusters are formed through recursive partitioning guided by a Gaussian-based entropy function that maximizes information gain. This enables the estimation of empirical distributions of efficiency scores, from which we derive confidence intervals for the Malmquist Index and its components—namely, catch-up and frontier-shift effects. The proposed method is validated using both simulated and real-world datasets, demonstrating its effectiveness in productivity analysis under stochastic environments.

5 - Efficient Statistical Tests for Personalization Effects

Zhaoqi Li, Stanford University, Stanford, CA, United States, Emma Brunskill

Personalized decision rules can outperform a single "one-size-fits-all" action, but they are worth learning only when personalization truly improves outcomes. In data from many real-world settings, one treatment dominates all other actions, and gathering the covariates needed to personalize can be costly. In these cases, it is of interest to test whether it is worth spending the effort to learn a personalized best policy, or a "one-size-fit-all" action suffices. To determine when personalization is justified, we propose a semiparametrically efficient test for the existence of the personalization effect, defined as the gain of the optimal context-dependent policy over the single best overall action. Our procedure uses a three-fold cross-fitted AIPW estimator that exploits the overlap between personalized and non-personalized recommendations and attains the semiparametric efficiency bound. We also use a repeated-split procedure that tightens finite-sample performance. We conduct simulations to show that the type-I error is controlled even when policy learning is imperfect. Our approach achieves a lower Type-I error when there is no personalization effect and a high power when there is personalization effect compared to the baseline. We also run real data experiments from various domains, including Jester jokes, MOOC interventions, and Job Corps earnings, that show the versatility of our algorithm in different settings. Together, these results provide an efficient and data-driven answer to the question "Is

personalization worth it?" and offer a ready-to-use toolkit for experimental design and policy evaluation across domains where individualized decisions may or may not pay off.

6 - High-Dimensional Statistical Process Control via Manifold Learning and Fitting

Ismail Tas, The Pennsylvania State University, State College, PA, United States, Enrique del Castillo

In modern industrial settings, processes often generate high-dimensional, temporally correlated data, challenging traditional Statistical Process Control (SPC) techniques. To address these limitations, we propose two novel Phase II SPC frameworks that address the geometry of high dimensional data under a manifold assumption. The first approach fits a smooth manifold in the ambient space and monitors deviations using a distribution-free univariate control chart. The second approach embeds high-dimensional observations into a lower-dimensional space and performs monitoring there using a distribution-free control chart. Both methods provide a controllable in-control average run length and demonstrate strong fault detection performance. We validate our approaches on the Tennessee Eastman Process Simulator, demonstrating their effectiveness in scenarios where traditional methods fail.

7 - Control Charts for Queues with Vacations

George Mytalas, Athens University Economics Business, Hlioupoli, Greece, Dimitris Zissis

In queues with vacations, the system benefits from vacation time to perform minor work as maintenance. Different policies are applied such as "Multiple", "Single" or "MAV." A well-designed control policy has a relative influence on the waiting time of customers as well as the service quality (QoS) of the system. For queues controlling the duration of vacations, the expected number of customers in the system is an important characteristic. We develop control charts for monitoring the expected system size in the M/M/1 queueing system with exponential vacation durations. Two methods of system parameters estimation applied for different vacation policies for control limit calculations are provided. Finally, we present an application example for illustrative purposes.

MC64

Building B Level 4 B404

AI and Worker Engagement

Invited Session

Behavioral Operations Management

Chair: Michelle Kinch, Tuck School of Business at Dartmouth College, Hanover, NH, United States

1 - Learning on the GO: How Gig Workers Learn with Algorithmic Recommendations

Park Sinchaisri, University of California Berkeley, Berkeley, CA, United States

Gig platforms increasingly rely on AI recommendation systems to guide workers in choosing tasks, yet we know little about how these systems shape decision-making and skill development over time. This study examines how gig workers learn, how their reliance on AI evolves, and what design principles can help recommendation systems support both short-term efficiency and long-term growth. Using detailed data from a retail delivery platform, we find early benefits from following AI guidance, but also evidence that experienced workers perform best when developing their own strategies. These insights motivate the design of adaptive, learning-aware AI systems that align with worker development and improve decision-making in dynamic environments.

2 - Worker Reactions to (Fair) Algorithms

Clare Snyder, NYU Stern, New York, NY, United States, Samantha Keppler, Stephen Leider

Despite the growing emphasis on fair algorithm design in ML research, human decision-makers remain key mediators of algorithmic outcomes. When people can deviate from and disregard algorithms' advice, it is possible they will do so in a way that undermines fairness objectives—just as algorithm aversion causes people to deviate from algorithms in ways that undermine their accuracy. In this paper, we explore how human decision authorities, such as loan officers, interact with fair algorithms, with a view towards understanding how their behavior affects fairness and accuracy objectives. With an online experiment (N=476), we compare subjects' decision-making processes across four algorithm treatments: an accuracy-only algorithm, a fair algorithm with an aggressive-threshold policy, a fair algorithm with a conservative-threshold policy, or no algorithm at all. We find participants under-rely on all algorithms, but use the fair algorithms' advice at particularly low rates. More critically, subjects' particular deviations from the fair algorithms result in lower fairness—measured by equality of opportunity—than participants without any algorithm or the fair algorithms themselves. This study contributes to an understanding of the way human behavior interferes with the effectiveness of fair algorithms and offers potential implications for improving human-algorithm collaboration in practical settings.

3 - Multistage Monitoring

Yue Li, Beijing Jiaotong University, Beijing, China, People's Republic of, Shuaikun Hou, Meng Li, Chao Liang

This paper explores the impact of algorithmic surveillance on task performance in a multistage sales process. We partner with a leading automobile company in Asia that has implemented algorithmic surveillance in select stores to provide real-time performance feedback to its call agents. Using a difference-in-differences approach, we examine the impact of algorithmic surveillance on agents' task initiation and task completion behaviors. Our analysis reveals that the introduction of algorithmic surveillance significantly reduces lead time—the duration between the assignment of an outbound call task and the initiation of the call—during the task initiation stage, without compromising task completion. In fact, we observe an increase in the conversion rate—the percentage of contacted customers who visit the store—indicating improved performance in the task completion stage as well. Moreover, our study explores the broader implications of algorithmic surveillance on workplace fairness, focusing on how the composition of different employee groups influences its effects on task performance. Our analysis reveals that not only do different employee groups (based on dimensions of gender, tenure, and position) exhibit varying responses, but their performance also differs depending on the task complexity. Our work highlights the important role of algorithmic surveillance in improving task performance in multistage processes and provides valuable insights for the development of a more equitable work environment.

4 - Negative Effects of Positive Real-TIME Feedback from AI

Michelle Kinch, Tuck School of Business at Dartmouth College, Hanover, NH, United States

Companies often rely on standard work protocols to increase efficiency in service delivery, but worker compliance remains a persistent challenge. This study examines whether real-time feedback from artificial intelligence ("AI-RTF") improves script adherence, task completion times and job satisfaction in customer contact center settings, while also assessing how its psychological impact on workers affects outcomes.

5 - Explaining Inconsistent Predictions

Zezen He, Simon Business School, University of Rochester, Rochester, NY, United States, Yaron Shaposhnik

In various machine learning applications, such as wait time prediction for customers in service, the model is repeatedly queried for predictions over time. This often occurs in response to changes in the dynamic characteristics of the user or the environment. Due to randomness, in many environments, predictions may change in ways that could seem unexpected or unreasonable to the user. For instance, predictions might oscillate between higher and lower values over time, a phenomenon we refer to as *intertemporal inconsistency*. For example, it is not uncommon for a customer to receive a short initial predicted wait time that later turns out to be a longer delay.

Motivated by research in psychology on adverse reactions to inconsistent behavior, we introduce the notion of *intertemporal explanations*, a novel form of explanation for model predictions that specifically aims to inform the user about the reasons behind intertemporal inconsistencies. We extend this method and further introduce the notion of *differential explanations* as information provided to users of machine learning models that explains why a model assigns different predicted values to two observations.

MC65

Building B Level 4 B405

AI and Future of Work

Invited Session

Artificial Intelligence

Chair: Zhiwei Qin, Amazon, San Jose, United States

Co-Chair: Xue Guo, Georgia State University, Atlanta, GA, 30303, United States

1 - How the Market Prices AI: Separating Adoption from Intention

John Sudeep Teli, Georgia State University, Atlanta, GA, United States, Arun Rai, Yu-Kai Lin

Artificial intelligence (AI) is an increasingly important driver of firm valuation, yet how the market prices AI remains incompletely understood. While firms frequently emphasize AI to enhance innovation and efficiency, the complexity and uncertainty surrounding AI make it difficult to assess its value. In this study, we investigate how the market prices AI by using data from 10-K filings. We distinguish firms' engagement in AI along two dimensions: AI Emphasis Current (AIEC), which reflects the extent to which a firm currently integrates AI into its business model, and AI Emphasis Planned (AIEP), which captures the firm's stated future intentions to adopt AI. Our results show that AIEC significantly predicts future stock returns, whereas AIEP does not, suggesting that the market underprices realized AI adoption but discounts forward-looking AI plans. We also examine changes in AIEC and AIEP as proxies for increases in current and planned AI engagement. We find that the market has no reaction to increases in AIEC in the short term but has a negative reaction to increases in AIEP. This implies that while the market overlooks new developments in AI adoption initially, it remains skeptical of newly stated intentions to adopt AI, possibly due to perceived implementation risk. Our study contributes to the literature on the economic value and risks of AI by focusing how different forms of AI engagement are valued, and to the mispricing literature by showing how investor pricing of AI-related information deviates from standard pricing fundamentals.

2 - Probing Personality with AI: Evaluating LLMs as Interviewers in Automated Hiring Systems

Rachit Kamdar, University of Maryland, College Park, MD, United States, Balaji Padmanabhan, Siva Viswanathan

The automation of hiring processes is increasingly central to organizational strategies aimed at enhancing recruitment efficiency and reducing costs. While resume screening remains the most widely used automation tool in recruitment, organizations are increasingly exploring ways to automate the interview process to further enhance hiring efficiency and reduce costs. Large language models (LLMs) offer a promising avenue for automating complex interviews through human-like conversational capabilities. However, their effectiveness in accurately assessing job-relevant characteristics, particularly personality traits that are critical to job performance remains under explored. This study investigates the effectiveness of LLM-conducted dynamic interviews, contrasting confrontational and non-confrontational interviewing styles, with traditional static question-based interviews in predicting self-reported personality traits. Drawing on established interviewing theories, we develop a novel interview design framework and evaluate it through a randomized controlled trial. Our findings reveal that LLM-driven dynamic interviews yield higher predictive accuracy of personality traits compared to static formats. Notably, confrontational LLM interviewers demonstrate superior accuracy in detecting negative traits such as neuroticism and disagreeableness. These insights have significant implications for the design of automated recruitment systems, highlighting the trade-offs between interview style, automation, and the accuracy of candidate assessments.

3 - Bandit-Based Offer Reshaping for Unassigned Tasks in Crowdsourcing Platforms

Young Yoon, Walmart, Sunnyvale, CA, United States, Yu Wang, Ibrahim El Shar, Hengchao Wang, Thomas Stiefel, Sherry Lijie Wan, Mingang Fu, Samuel Eldersveld

A key challenge faced by task requesters on crowdsourcing platforms is determining how to reshape task offers when they remain unassigned. This problem is particularly prevalent across many delivery platforms, where demands are predetermined, and task requesters need to make the unassigned offers more attractive to ensure timely acceptance. For example, task requesters can add subtasks with different prices to make the task more appealing. While prior research has focused on the initial matching of tasks with workers, studies on reshaping task offers are sparse.

We propose a bandit-based offer reshaping mechanism that utilizes a reverse-Dutch auction method to gradually modify task offers, such as by incrementally adding subtasks. The rate of adjustment (i.e., reshaping pattern) is determined using a novel pattern algorithm with parameters optimized through a multi-objective multi-armed bandit (MOMAB) framework. Additionally, we present an end-to-end system that integrates batched offline learning with online inference.

There are two additional challenges to our problem. First, task requesters commonly aim to improve a success metric while maintaining a guardrail metric within defined bounds. The challenge arises from the need to adapt to dynamic priorities when guardrail metrics fall outside of the bounds. We address this by using lexicographically scalarized expected rewards to prioritize objectives. Second, estimating expected rewards for batched learning is challenging due to the limited number of unassigned tasks per batch. We propose a new confidence-based learning method that dynamically adjusts learning rates.

Through offline evaluation, we provide empirical evidence that our approach achieves a regret of $O(\log^2(T))$.

MC66

Building B Level 4 B406

Teaching the Future: AI, Ethics, and Inclusion in the Next Generation of Education

Panel Session

INFORMS Committee on Teaching and Learning

Co-Chair: Gabriela Gongora-Svartzman, Khoury College, Northeastern University, Miami, FL, United States

1 - Panelist

Sourav Chatterjee, University of North Texas, Denton, TX, United States

2 - Panelist

Holly Wiberg, Carnegie Mellon University, Pittsburgh, PA, United States

3 - Panelist

Gabriela Gongora-Svartzman, Northeastern University, Miami, FL, United States

4 - Panelist

Angelika Leskovskaya, SMU Cox School of Business, Dallas, TX, United States

MC67

Building B Level 4 B407

Recent Advances in Spatiotemporal Data Mining

Invited Session

Data Mining

Chair: Adam Meyers, University of Miami, Miami, FL, United States

1 - Hybrid Diffusion-Transformer Approach for Multivariate TIME Series Imputation**Ramiro Valdés, University of Miami, Miami, FL, United States, Adam Meyers**

Many real-world spatiotemporal data contain gaps when sensors fail or communication issues occur. Spatiotemporal imputation focuses on reconstructing missing values by leveraging patterns that repeat across both space and time. In recent years, deep learning approaches have emerged as powerful tools for spatiotemporal imputation, consistently outperforming traditional methods. Although recent work has improved performance, two persistent gaps remain: (i) diffusion-based methods model uncertainty well but are confined to local temporal contexts and fixed spatial graphs, and (ii) transformer-style architectures capture long-range dependencies yet handle missing-value uncertainty only heuristically. We propose a hybrid method that addresses the limitations of each individual approach, namely, the Hybrid Diffusion Transformer (HDT). The diffusion component treats each time series as a noisy signal that can be gradually reconstructed, giving the model a way to handle uncertainty and to learn to remove the noise step by step. The transformer component adds an attention mechanism, enabling the model to simultaneously learn spatial correlations and non-stationary temporal patterns. During inference, the reverse diffusion trajectory, guided by the transformer component, iteratively reconstructs missing values while producing confidence intervals. By integrating these parts in a single, end-to-end framework, our approach offers a robust and interpretable solution for multivariate time series imputation. We evaluate HDT on varied traffic and physiological datasets chosen to test long-range and spatiotemporal dependencies. Model validation covers a range of missing-value patterns and compares mean absolute error (MAE) and root mean square error (RMSE) against state-of-the-art baseline methods.

2 - A Statistical Model for Monitoring Convolution-Generated Spatio-Temporal Gaussian Processes**Yutong Zhang, Georgia Institute of Technology, Atlanta, GA, United States, Xiao Liu**

A large class of Gaussian spatio-temporal processes can be constructed through a sequence of convolution operations in time and space. For example, stochastic advection-diffusion processes with a Gaussian noise, heat equations with a random force driven by a Gaussian process, etc. For many of these processes, one is interested in monitoring the first-order time derivative which plays a critical role in change point detection of such processes. In this paper, we propose a statistical model that is effective in monitoring the first-order time derivative of convolution-generated Gaussian space-time processes. An infinite-dimensional stochastic differential equation (SDE) representation is firstly obtained. Then, a finite-dimensional approximation, based on the Gelerkin projection, is proposed to yield a computationally-efficient state-space representation. Numerical investigations are performed based on both a simulation dataset and a real dataset of the Aerosol Optical Depth during 2025 Los Angeles wildfires. Comparison studies show that the proposed approach provides significant advantages over existing approaches in monitoring convolution-generated space-time Gaussian processes.

3 - Topology-Aware Conformal Prediction for Stream Networks**Fangxin Wang, University of Illinois Chicago, Chicago, IL, United States**

Stream networks, a unique class of spatiotemporal graphs, exhibit complex directional flow constraints and evolving dependencies, making uncertainty quantification a critical yet challenging task. Traditional conformal prediction methods struggle in this setting due to the need for joint predictions across multiple interdependent locations and the intricate spatio-temporal dependencies inherent in stream networks. Existing approaches either neglect dependencies, leading to overly conservative predictions, or rely solely on data-driven estimations, failing to capture the rich topological structure of the network. To address these challenges, we propose Spatio-Temporal Adaptive Conformal Inference (STACI), a novel framework that integrates network topology and temporal dynamics into the conformal prediction framework. STACI introduces a topology-aware nonconformity score that respects directional flow constraints and dynamically adjusts prediction sets to account for temporal distributional shifts. We provide theoretical guarantees on the validity of our approach and demonstrate its superior performance on both synthetic and real-world datasets. Our results show that STACI effectively balances prediction efficiency and coverage, outperforming existing conformal prediction methods for stream networks.

4 - Geometry-Aware Active Learning of Spatiotemporal Dynamic Systems**Xizhuo Zhang, University of Tennessee, Knoxville, TN, United States**

Rapid developments in advanced sensing and imaging have significantly enhanced information visibility, opening opportunities for predictive modeling of complex dynamic systems. However, sensing signals acquired from such complex systems are often distributed across 3D geometries and rapidly evolving over time, posing significant challenges in spatiotemporal predictive modeling. This paper proposes a geometry-aware active learning framework for modeling spatiotemporal dynamic systems. Specifically, we propose a geometry-aware spatiotemporal Gaussian Process (G-ST-GP) to effectively integrate the temporal correlations and geometric manifold features for reliable prediction of high-dimensional dynamic behaviors. In addition, we develop an adaptive active learning strategy to strategically identify informative spatial locations for data collection and further maximize the prediction accuracy. This strategy achieves the adaptive trade-off between the prediction uncertainty in the G-ST-GP model and the space-filling design guided by the geodesic distance across the 3D geometry. We implement the proposed framework to model the spatiotemporal electrodynamics in a 3D heart geometry. Numerical experiments show that our framework outperforms traditional methods lacking the mechanism of geometric information incorporation or effective data collection.

5 - Multi-Graph Structure Learning for Multivariate TIME Series Anomaly Detection**Jaeyeon Lee, KAIST, Daejeon, Korea, Republic of, Heeyoung Kim**

Industrial processes often generate complex, interdependent time-series data from multiple sensors across multiple stages, forming complex dependencies among variables and process stages. Effective monitoring and timely anomaly detection of these time series through multivariate time series anomaly detection (MTAD) are crucial for preventing failures and maintaining system reliability. Graph neural networks (GNNs) have advanced MTAD by leveraging data-driven graphs to model complex dependencies among variables, effectively capturing relational structures within multivariate time series to enhance anomaly detection performance. We propose a novel GNN-based MTAD method that constructs and integrates multiple complementary graphs to better represent inter-variable relationships. Comprehensive experiments on two real-world, multi-stage industrial datasets demonstrate that our method outperforms state-of-the-art baselines in anomaly detection performance.

6 - Adaptive Sampling for Concept Drift Detection**Junghye Pyeon, Georgia Institute of Technology, Alpharetta, GA, United States, Davide Cacciarelli, Kamran Paynabar**

Concept drift and label scarcity are two critical challenges limiting the robustness of predictive models in dynamic industrial environments. Existing drift detection methods often assume global shifts and rely on dense supervision, making them ill-suited for regression tasks with local drifts and limited labels. This paper proposes an adaptive sampling framework that combines residual-based exploration and exploitation with EWMA monitoring to efficiently detect local concept drift under label constraints. Empirical results on synthetic benchmarks and a real-world electricity market case study demonstrate superior performance in label efficiency and drift detection accuracy.

MC68

Building B Level 4 B408

AI-Driven Methods for Clinical Decision Support

Invited Session

Data Mining

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

1 - Holistic AI in Medicine: Improving Performance and Explainability**George Margaritis, Massachusetts Institute of Technology (MIT), Cambridge, MA, United States, Periklis Petridis, Vasiliki Stoumpou, Dimitris Bertsimas**

Deployment of Artificial Intelligence (AI) solutions in real-world clinical settings faces critical challenges in ensuring both prediction accuracy and explainability. Extracting useful and interpretable features from the overwhelming volumes of patient data remains a largely unsolved problem. Existing approaches typically rely on averaging embeddings over multiple time instances and data elements (e.g. clinical notes), often degrading predictive performance and obscuring clinical meaning. We

present xHAIM (Explainable Holistic Artificial Intelligence in Medicine), a novel framework that attempts to address these limitations through a generalizable 4-step approach. xHAIM automatically identifies task-relevant patient data across multiple modalities, generates comprehensive task-specific and patient-specific summaries, uses these summaries to enhance predictive performance, and provides clinically relevant explanations by combining predictions with existing medical knowledge. Our evaluation using the MIMIC-IV dataset demonstrates significant performance improvements over the baseline, with AUC increases that reach as much as 11% in chest pathology prediction tasks. We also show that in the majority of cases, the system is able to correctly justify model predictions by citing relevant sections of the patient's Electronic Health Record, thus paving the way for an accurate and interpretable clinical decision support system.

2 - Automated Differential Diagnosis: A Pomdp Framework for Optimal Diagnostic Test Selection in the Emergency Department

Karl Zhu, Massachusetts Institute of Technology, Cambridge, MA, United States, Yubing Cui, Matthew Peroni, Vasiliki Stoumpou, Cindy Wang, Dimitris Bertsimas

We present a POMDP model for automating diagnostic test selection in the Emergency Department (ED), where clinicians must quickly and accurately perform differential diagnosis. Using real electronic health records from Hartford HealthCare—one of the largest hospital systems in Connecticut—we initialize a belief over 25 Major Diagnostic Categories using an XGBoost model trained on demographics, vitals, and chief complaints. This belief is refined through the results of diagnostic lab and imaging tests (e.g., metabolic panels, urinalysis, ultrasound). The objective minimizes both test costs and final diagnostic uncertainty, measured by entropy. Solving the POMDP exactly for each patient and simulating synthetic test outcomes, we show that the optimal policy significantly improves diagnostic accuracy—demonstrating its potential as a clinical decision support tool.

3 - Road 3: Mitigating Observed and Unobserved Confounding in Observational Data

Vasiliki Stoumpou, Massachusetts Institute of Technology, Cambridge, MA, United States, Georgios Antonios Margonis, Dimitris Bertsimas

Randomized clinical trials are widely regarded as the gold standard for assessing treatment effects in medical settings. In contrast, observational data lack randomization, leading to imbalances between treatment arms due to confounding factors. While many causal inference methods attempt to adjust for observed confounding, they often operate under the assumption that unobserved confounding is absent. In this work, we challenge that assumption by proposing a novel method to estimate the portion of patient outcomes that cannot be explained by observed variables—effectively introducing a proxy for unobserved confounding. We then leverage this proxy, alongside the known medical covariates, to construct a balanced cohort that better emulates a randomized trial population. We evaluate our approach across four datasets—two purely observational and two randomized trials with synthetically introduced confounding—and show that it consistently outperforms standard methods that account for observed confounding only. Our method offers a promising direction for enabling personalized treatment decisions in contexts where treatment effects are not known a priori.

4 - Detection, Localization and Progression Prediction of Subdural Hematoma

Rohan Kumar, Massachusetts Institute of Technology, Cambridge, MA, United States

A subdural hematoma is a type of intracranial bleed that occurs on the surface of the brain. Aging, trauma in the head, use of anticoagulants and alcohol, and coagulopathies have been linked to the occurrence of subdural hemorrhages. Subacute/chronic subdural hematoma is expected to emerge as a predominant neurosurgical concern in the United States by 2030, primarily attributed to increasing life expectancy and the expanding utilization of antiplatelet and anticoagulation medications. The present work evaluates the efficacy of machine learning (ML) approaches used to predict the existence, localization, and progression of chronic subdural hematoma. We perform these tasks using demographics, labs, medications and image data and utilizing deep learning architectures to perform classification and segmentation of the images, to detect the area of the hematoma. Then, we employ the segmented scans and hematoma existence prediction results to understand the prognosis of the condition with or without clinical intervention. Our aim is to provide clinicians with a robust AI tool that provides them with information regarding the probability of the existence, location, and progression of a hematoma.

MC69

Building B Level 4 B409

Innovation in Services

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Abhishek Deshmane, Georgia Institute of Technology, Atlanta, GA, United States

1 - New Technology Adoption and Peer Effects Among Service Workers: Evidence from Aadhaar Enabled Payment System

Deepanshi Bhardwaj, University College London, London, United Kingdom, Bilal Gokpinar, Ashwini Chhatre

Given the increasing pace of new technology introductions and associated adoption challenges experienced by service workers, our study investigates whether and to what extent peers affect service workers' performance during new technology adoption. Using a granular transaction level dataset of bank-service agents in India, and examining the introduction of a new digital payment system (Aadhaar Enabled Payment System – AePS), we find that at the onset of AePS, agents with peers exhibit a 6% lower failure rate of transactions in comparison to agents with no peers. We also find that more experienced peers provide stronger peer effects than less experienced peers. In addition, competition hinders peer effects in that having peers who compete for the same customer pool or peers who are competitive based on their job nature (full-time in comparison to part-time) significantly weakens benefits gained from having peers. Our study provides novel and detailed empirical evidence of peer effects at the onset of a new technology introduction. Our findings contribute to the literature on innovation and technology management by extending the scope of peer effects from day-to-day operations to the critical phase of new technology adoption, and offer important insights into organization and management of service workers during this critical period.

2 - Small and Medium Enterprises in Open Innovation

Gaoyu Xie, George Washington University, Washington, DC, United States, Pascale Crama, Janne Kettunen

We develop a game-theoretic model to examine the innovation strategies of small and medium-sized enterprises (SMEs), focusing on two key frictions that differentiate them from large firms: risk aversion and the cost of capital. These frictions contribute to both innovation and launch inefficiencies, ultimately suppressing SMEs' innovativeness. Our model explores whether establishing an open innovation (OI) ecosystem—where SMEs collaborate with large supplier firms—can enhance innovation outcomes. We find that involving suppliers in the innovation process reduces innovation inefficiency, although launch inefficiency may persist when fixed costs are low. Participation in OI is particularly beneficial under high market uncertainty but provides limited value when both fixed costs and technical uncertainty are high.

3 - Double-Edged Sword: Impact of AI on Production of Knowledge

Hyunwoo Park, Seoul National University, Seoul, Korea, Republic of, Seokbeom Kwon, Hyun Ju Jung

This study examines how AI integration reshapes research performance and collaborative dynamics among scientists facing potential displacement by AI tools. Using DeepMind AlphaFold's 2018 entry into the CASP13 protein structure prediction competition as an exogenous shock, we employ a difference-in-differences design on scientists' publication and citation panel data. Our research investigates differential impacts across career stages, examining how senior and junior scientists' productivity and impact metrics respond to AI introduction. Through co-authorship network analysis, we explore how scientists reposition themselves within collaborative structures following AI disruption. We also analyze whether strategic AI adoption can mitigate potential adverse effects on research impact. This work contributes to our understanding of AI's multifaceted influence on scientific communities and knowledge production institutions. Our findings provide insights into career-stage-dependent effects of AI augmentation on scientific practice, informing both policy and individual career strategies in an increasingly AI-influenced research landscape.

4 - Vincent's Talk

Vincent Peters, Tilburg University, Tilburg, Netherlands

Monday PC

Building A Level 1 Exhibit Hall A3

Monday Poster Competition Session

Poster Session

Contributed

1 - A Hyperheuristic Hybridized with Exact Optimization for the Berth Allocation and Scheduling Decision Problem

Bokang Li, Florida A&M University-Florida State University, Tallahassee, FL, United States, Fatemeh Shekoochi, Payam Afkhami, Maxim Dulebenets

Berth allocation and scheduling at marine container terminals holds critical significance for optimizing port operations and ensuring efficient maritime logistics, yet it is exceptionally challenging due to its operational complexity and practical constraints. This study presents a mixed-integer linear programming model for the dynamic discrete berth allocation and scheduling problem with the objective to minimize the total turnaround cost and develop an efficient service schedule for each incoming vessel with a feasible service order and berthing position. A novel Hyperheuristic Hybridized with Exact Optimization (HHEO) is developed to explicitly solve the challenging decision problem studied herein. The HHEO algorithm is designed to dynamically select and apply different genetic operators based on their actual performance. Furthermore, intelligent exact optimization procedures specific to the domain of berth allocation and scheduling are periodically employed within the HHEO framework to improve solution quality and facilitate search for high-quality solutions. Detailed computational experiments are carried out to prove the superiority of the HHEO algorithm against solution methods inspired by exact optimization along with some well-known metaheuristic algorithms, demonstrating its efficiency and applicability for real-life berth planning at marine container terminals. Moreover, the experiments clearly demonstrate that incorporating a hyperheuristic framework along with problem-specific hybridization methods is critical for improving the exploratory and exploitative capabilities of the developed HHEO algorithm. Last but not least, valuable practical insights are uncovered through the proposed approach, offering reliable solutions that can support effective port management under different scenarios of berthing availability, vessel arrival intensity, and delayed vessel departure penalties.

2 - Multiphysics-Informed Neural Network for Simulating Electromechanics of Atrial Fibrillation

Weixuan Dong, The University of Oklahoma, Norman, OK, United States, Yifu Li, Stavros Stavrakis, Rui Zhu

Realistic patient-specific simulations are crucial for treating atrial fibrillation (AF), a cardiac arrhythmia driven by multi-physical processes. Current research mainly focuses on single-physics simulations due to computational complexity. To address this, we develop a Multiphysics-informed neural network that captures AF's complex dynamics, bringing simulations closer to real human hearts.

3 - AI Impact on Digital Marketing: Detecting Behavioral Changes

Melody Han, St. John's School, Houston, TX, United States, Zijun Zhan, Tony Cui

Using Fiverr data, we examined the impact of AI adoption on client retention and service demand using regression and CUSUM methods. Our findings reveal significant behavioral shifts triggered by AI, offering valuable insights into platform strategy, customer lifecycle management, and the timing of technology-induced disruptions in digital marketplaces. These insights can guide managers in optimizing service offerings, enhancing customer engagement, and maintaining competitive advantage in a rapidly evolving market.

4 - Multi-Task Learning-Based Shared Encoder Framework for Robust Anomaly Detection

wonjae Kim, Department of Industrial and Management Engineering, Incheon National University, Incheon, Korea, Republic of, Jaehong Yu

Recently, the importance of anomaly detection for early detection of equipment abnormalities in industrial sites has been highlighted. However, existing anomaly detection methods do not reflect the distribution differences between normal data collected in various operating environments, resulting in a problem of reduced anomaly detection performance. In this study, we propose a methodology to robustly detect anomalies in various operating environments by converting signal data into spectrograms, forming a cluster structure of normal data in latent space using deep clustering-based learning, and calculating a novelty score based on this.

5 - Integrated Multi-Objective Optimization and Spatial Analysis for the Site Selection of Levee Crossing Facilities

Tsun-Yuan Ku, Department of Geomatics, National Cheng Kung University, Tainan, Taiwan, Febrian Fitryanik Susanta, PEI-FEN KUO, Chung-Wei Shen, Ching-Jui Lai, Nan-You Chen

In many large riverside cities, levee systems serve dual roles in flood control and urban recreation by enabling parks and jogging paths along and beyond embankments. However, pedestrian access to these off-levee spaces often depends on crossing infrastructure, such as stairs and ramps, originally intended for maintenance rather than public use. These facilities often lack accessibility compliance, posing barriers for elderly and mobility-impaired users. Moreover, the spatial distribution of these crossings is frequently uneven and poorly integrated with pedestrian networks.

This study focuses on the Sanchong and Luzhou districts of New Taipei City, Taiwan, an area of 560,000 residents within 23 km², enclosed by a circular embankment over 16 km long and currently served by 103 levee-crossing facilities. A smart site selection framework is proposed to prioritize facility upgrades under budget constraints, aiming to improve walkability and spatial equity. The model addresses three objectives: (1) maximizing accessibility by meeting pedestrian demand, (2) minimizing sidewalk retrofit costs in high-demand areas, and (3) minimizing reconstruction costs of crossing facilities. A spatial dispersion mechanism discourages clustering and improves overall network efficiency.

To address the complex, multi-objective nature of the levee crossing site selection problem, we adopt two optimization approaches: the Non-dominated Sorting Genetic Algorithm III (NSGA-III) and a Quadratic Unconstrained Binary Optimization (QUBO) formulation. Both methods generate a well-distributed Pareto front while ensuring solution diversity. The solution sets are compared in terms of quality, spatial feasibility, and trade-off structure, providing effective solutions to this site facility location problem.

6 - Overconfident Pitches, Vulnerable Investors: the Role of Financial Literacy in Crowdfunding and the Need for Consumer Protection Nudges

Andres Riofrio-Lara, Ivey Business School, London, ON, Canada, Elena Fumagalli, Giovanni Visentin, Fabiana Penas

Crowdfunding democratizes access to capital for entrepreneurs, but it also exposes consumers to heightened financial risk. Platforms like Kickstarter face persistent concerns regarding fraud and non-delivery, with nearly 18% of the funds raised estimated to stem from deception campaigns. We investigate how founder overconfidence, manifested as overestimation, overplacement and overprecision, affects consumer pledging behavior.

We develop a linguistic manipulation of a crowdfunding pitch and examine its interaction with consumers' financial literacy. In our initial experiment (n = 200), participants were randomly assigned to view either a neutral or an overconfident pitch for a 3D printer project, then selected investment alternatives mirroring Kickstarter's structure. Financial literacy and risk preferences were subsequently assessed using validated scales. Results show that exposure to an overconfident pitch significantly increased pledging among financially illiterate consumers, while financially literate participants exhibited greater skepticism and lower investment levels. These effects remained robust across multiple model specifications, even after controlling for investor overconfidence, domain knowledge, and prior crowdfunding experience. Additionally, we identify an interaction between risk preferences and the treatment that also shapes the investment decision.

These findings highlight the phenomenon of overconfidence contagion in consumer decision making, underlining a mechanism by which cognitive biases can exacerbate financial vulnerability. Our work contributes to the literature on behavioral finance, fintech consumer protection and decision hygiene. Future research will explore the role of overconfidence cues and the potential of decision hygiene nudges to mitigate cognitive bias and enhance investor safeguards in crowdfunding environments.

7 - An Integrated Web Tool for Nutrient Optimization and Watershed Model Selection in Harmful Algal Bloom Mitigation

Ashim Khanal, University of South Florida, Tampa, FL, United States, Hadi Charkhgard, Qiong Zhang, Mauricio E. Arias, David Kaplan

Harmful algal blooms (HABs) are intensifying in frequency and severity, posing significant threats to public health, aquatic ecosystems, and the economy. Excessive nutrient loading—primarily phosphorus (P) and nitrogen (N)—is a major driver of HABs, often stemming from agricultural and urban runoff. While Best Management Practices (BMPs) offer a promising solution to reduce nutrient loads, selecting the most effective BMPs and their optimal placement within a watershed remains a complex challenge.

This poster presents a user-friendly, web-based decision support tool that integrates watershed modeling and optimization to aid in HAB mitigation. The tool features a model selector that helps users identify the most suitable hydrological model—such as WAM or SWAT—based on watershed characteristics. It includes backend automation that transforms model outputs into the format required by AquaNutriOpt, a surrogate-based optimization framework that identifies cost-effective BMP strategies under uncertainty.

The web tool allows users to perform single or bi-objective optimization for P and N across multiple future scenarios. It includes interactive visualization, guided tutorials, and downloadable documentation to enhance accessibility for diverse stakeholders. We have demonstrated the Optimization engine's effectiveness through a case study with Watershed Assessment Model (WAM) as primary hydrological model and watersheds of Lake Okeechobee as study area. This integrated platform empowers water managers, researchers, and policymakers with a streamlined, science-based approach for controlling nutrient pollution and mitigating HABs.

8 - Estimating Potential Future Profitability of Battery Storage Using Electricity Market Price Decomposition

Julia Frohmann, Stanford University, Stanford, CA, United States

Decarbonizing the U.S. electricity grid requires large-scale deployment of battery energy storage systems (BESS), spanning durations from 4 to 100+ hours. However, the pace of deployment is constrained by uncertain project profitability, particularly as revenue opportunities decline due to increasing wholesale market saturation (a trend already impacting ancillary services in CAISO and ERCOT). To address this, we constructed a future electricity price generator to assess potential scenarios of storage revenues under evolving energy and ancillary services market conditions.

Our approach employs Discrete Fourier Transform to decompose historical data and generate potential future price time series. The generator allows user-defined modulation of duck curve depth, underlying trends, and stochastic peak events, capturing typical diurnal and seasonal patterns well. Coupled with the battery dispatch optimization model DER-VET, we simulate BESS operations for 41 CAISO nodes, 9 future scenarios, and storage durations from 1 to 12 hours.

Revenue optimization results from synthetic price data are 6.7% lower than real price results, yielding a conservative estimate potential revenues. When comparing different generated potential future scenarios, we find that revenues can drop by 23.3% if both the energy arbitrage potential and the ancillary services prices decrease compared to 2023 revenue, making these projects much less attractive to developers. Also, BESS beyond 6h duration are uneconomical in the current market. These findings underscore the urgent need to develop storage-adapted market and contract mechanisms to compensate long(er) duration storage providers and enable cost-optimal buildout.

9 - Learning-Based Cpu Burst Prediction for Efficient Scheduling in Operating Systems

Md Nazmus Shakib, University of Alabama in Huntsville, Huntsville, AL, United States, Rahul Bhadani

In high-performance computing environments, scheduling decisions rely heavily on estimating how long a process will occupy the CPU. Conventional estimation techniques often fall short under varying workloads, limiting their effectiveness in modern, multi-user systems. This work introduces a data-driven framework that uses machine learning to predict process burst times using only features known at submission time. We evaluate several machine learning models, including decision trees, random forests, and multi-layer perceptrons, using real-world scheduling traces. Among them, the decision tree model offers the best trade-off between accuracy and speed. It achieves a mean absolute error (MAE) of 3514.09 and relative absolute error (RAE) of 12.64%, which is significantly lower than the exponential averaging baseline (MAE: 14833 ms, RAE: 53.34%), while keeping inference time per sample around 0.0033 milliseconds, making it suitable for real-time scheduling. Compared to exponential averaging, the learning-based models provide improved accuracy across a variety of workload conditions. The early availability of accurate predictions allows schedulers to make better decisions without introducing significant delay. This helps reduce wait times and improves overall CPU utilization. The study also includes model tuning and performance validation using 95% confidence intervals to ensure the results are consistent and reliable. By introducing a fast and accurate prediction method, this work supports efficient CPU scheduling strategies that respond well to real-world system demands.

10 - Decomposition-Based Framework for Optimizing Transit Stop and Route Placement to Enhance Accessibility in Transit Deserts

Md Gulam Kibria, Georgia Institute of Technology, Atlanta, GA, United States, Srinivas Peeta

Public transit offers a more sustainable and affordable alternative to personal vehicle travel, yet its adoption remains limited in many urban areas due to insufficient service coverage and poor connectivity. These challenges are particularly severe in transit deserts, where residents lack reasonable access to reliable transit options. This study addresses these challenges by developing an optimization framework to strategically expand transit stop locations and route connections with the goal of improving access to essential destinations such as jobs, education, healthcare, and grocery stores. It formulates a bi-objective optimization problem that identifies the optimal spatial configuration of transit stops and the connecting network of links. The two objectives are: (i) to maximize accessibility to key destinations across the service region, and (ii) to minimize the infrastructural and operational costs associated with new stop and link placements. To ensure computational efficiency for mid-sized and large-sized networks, a decomposition-based solution algorithm is proposed that iteratively solves reduced subproblems for stop selection and route connectivity while maintaining solution feasibility. Numerical experiments are conducted on both a synthetic grid network and a real-world network representing the City of Peachtree Corners, GA. Results show that the proposed approach significantly improves accessibility in transit deserts and lowers the Gini coefficient of accessibility distribution. The framework provides practical insights that can inform a range of stakeholders involved in planning, funding, and operating transit systems.

11 - Nonlinear Spatio-Temporal Run-to-Run Control: Leveraging Diffusion Models for Disturbance Management

Zihan Zhang, Georgia Institute of Technology, Atlanta, GA, United States, Lingchao Mao, Jianjun Shi, Kamran Paynabar

Run-to-run (R2R) control is widely used in advanced manufacturing processes. Conventional R2R control methods often assume that system disturbances exhibit linear structures or low-rank approximations in high-dimensional data, such as images or videos. While these assumptions facilitate implementation, they limit the effectiveness of R2R control in managing sophisticated manufacturing processes with intricate nonlinear disturbances. To address these challenges, we propose a novel nonlinear spatio-temporal R2R control framework that estimates system disturbances using a diffusion model and applies control actions to compensate for them. The proposed approach integrates an offline modeling phase to capture disturbance dynamics and an online control phase that dynamically adjusts actions to minimize deviations from the target system response, ensuring system stability and precision. Unlike conventional deterministic control strategies, the proposed method explicitly quantifies uncertainty, enabling more robust and informed decision-making. The effectiveness of this method is validated through simulation studies and a case study, demonstrating its adaptability in complex, high-dimensional environments.

12 - The Madness of Multiple Entries in March Madness

Jeff Sylvestre-Decary, University of Connecticut, Storrs, CT, United States, David Bergman, Carlos Cardonha, Jason Imbrogno, Andrea Lodi

This paper investigates the effectiveness of maximizing the expected value of the best performing entry in multi-entry betting pools for single-elimination tournaments, with a particular focus on March Madness. In such betting pools, participants select winners for each game, and their score is a weighted sum of the correct selections. Due to the top heavy payout structures in these pools, we study whether maximizing the expected score of the highest-scoring entry, which has been successfully adopted in the sports betting literature, is a suitable modeling approach for March Madness betting pools. Unlike traditional methods that require modeling how other participants make their selections—an especially challenging task given that March Madness only occurs annually—this approach only requires win probability estimates for teams as input, and inherently provides a diversified position across multiple entries. We present an exact dynamic programming approach for calculating the expected maximum score of any fixed set of entries. Additionally, we explore the structural properties of this approach to develop several solution techniques. Using insights from one of our

algorithms, we design a simple yet effective heuristic which delivers high-quality results when tested against high-roller bettors in a real world March Madness betting pool. In particular, our results demonstrate that the best 100 entry solution identified by our approach had a 2.45% likelihood of winning a \$1 million prize and an expected profit of \$14,502.43 in a real-world betting pool, showing that the proposed heuristic is competitive even against the best professional bettors.

13 - Optimizing Pallet Utilization to Reduce Curing Costs in Precast Concrete Manufacturing

Anik Mazumder, Louisiana State University, Baton Rouge, LA, United States, Bhaba Sarker

Efficient curing of precast concrete (PC) components is vital for ensuring consistent material strength where meeting delivery deadlines and controlling production costs should also be considered. This study develops an optimization framework to find an optimal layout configuration to arrange PC components on pallets and to minimize batch or pallet cycles during the curing process. An mixed integer linear programming model has been formulated that assigns components to pallets under area and demand constraints, aiming to reduce the total number of curing batches and associated energy and labor costs. To address rising computational times as component sizes and demand volumes grow, a custom branch and bound heuristic is proposed to generate high-quality layouts rapidly. Computational experiments using industrial case study show that optimized curing can decrease cost by up to 20 % and cut CPU time by 30 % compared to baseline configurations. Sensitivity analysis highlights that beyond a certain packing density, additional consolidation yields diminishing returns, guiding practical batch-size decisions. By ensuring uniform curing conditions and reducing batch cycles, the approach not only lowers immediate production expenses but also contributes to long-term sustainability through reduced maintenance and repair needs. Future work will focus on identifying key variables for faster layout generation and integrating the model with real-time production scheduling systems.

14 - Llm-Enhanced, Data-Driven Personalized and Equitable Anesthesiologist Scheduling Under Uncertainty: a Joint Prediction and Optimization Approach

Anjali Jha, University of Maryland Baltimore County, Baltimore, MD, United States, Maxim Eckmann, Muhammad Ali Yousuf, Kai Sun

Anesthesiologist scheduling remains a challenge due to limited clinical resources and fluctuating demands. This complexity is especially acute in large academic anesthesiology departments as physicians balance responsibilities across multiple clinical sites with conflicting priorities. Further, scheduling must account for individuals' clinical and lifestyle preferences to ensure job satisfaction and well-being.

Traditional approaches, often based on statistical or rule-based optimization models, e.g., stochastic optimization or mixed-integer programming (MIP), rely on structured data and explicit domain knowledge. However, these methods often overlook unstructured information, e.g., free-text notes from routinely administered clinician well-being surveys and scheduling platforms. These notes may reveal implicit constraints (post-call recovery needs, early departures for family obligations) and underutilized clinical resources (a physician's willingness to work additional hours to address emergent needs).

Neglecting such information can lead to misaligned schedules, increased burnout, overlooked staffing flexibility, and suboptimal utilization of available resources. To bridge this gap, we propose a joint prediction and optimization framework that integrates MIP with LLM-based clinician preferences. The LLMs extract actionable preferences and implicit constraints from unstructured survey and scheduling notes, which then inform a multi-objective optimization program considering schedule consistency, fair workload distribution, and contractual full-time equivalent allocations.

To enhance reliability, a validation layer that treats LLM outputs as intermediate signals to reduce hallucination risks is implemented to ensure operational robustness. The proposed approach combines the interpretive strength of LLMs with the rigor of mathematical optimization to deliver a practical, data-driven solution that promotes clinician satisfaction and improves operational efficiency in complex healthcare settings.

15 - Decision Support Using a Bayesian Digital Twin for Personalized Medical Decision Making

Fangzhou Xie, Boston University, Boston, MA, United States, Yangyuechen Zhao

A real-time decision aid shows the personalized benefits and side effect probabilities for treatment options. It utilizes a digital twin virtual patient representation that integrates electronic health records and patient reported outcomes. A Bayesian framework (prediction model prior) continuously updates information (prediction model posterior) or alerts clinicians of inconsistencies.

16 - Decision-Aware Predictions for Right-Hand Side Parameters in Linear Programs

Jackson Forner, Southern Methodist University, Dallas, TX, United States, Miju Ahn, Harsha Gangammanavar

In this talk, we consider the following predict-then-optimize problem: (1) predict uncertain right-hand side parameters of linear programs (LPs) using associated context/feature data, and (2) solve the resulting deterministic LP. A natural issue that arises when making such predictions is that the resulting optimal solution (if one exists) may be infeasible under the true realization of the right-hand side parameters. With this in mind, we propose a class of *decision-aware prediction models* that learns the relationship between context and right-hand side parameters in a way that yields high feasibility of predicted solutions. These models contrast decision-unaware prediction models, such as linear or random forests regression models, which treat right-hand side parameters simply as target values, i.e., not as parameters of an optimization problem. To compare these two classes of models, we use a synthetic data experiment as well as two real-life network LP problems: a bioenergy supply chain problem with uncertain biofuel demand, and a beer distribution problem with uncertain consumer demand. We show that our proposed decision-aware models achieve higher out-of-sample feasibility when compared to the decision-unaware regression models, with only a mild loss in solution quality as measured by a gap in optimal objective function value.

17 - A Nash Equilibrium Model for Reverse Logistics for Electric Vehicle Batteries

MD Mohiuddin, University of Louisville, Louisville, KY, United States, Lihui Bai

The rapid growth of the electric vehicle (EV) market puts strain on the raw material supply for new EV batteries (EVBs). Recovering materials from end-of-life (EoL) EVBs offer a promising secondary source. This study develops a novel Nash equilibrium model to design a reverse logistics system for EVBs, consisting of three players, i.e., end-of-life EVB collectors, recyclers, and manufacturers. The equilibrium model describes each player's own profit maximization behavior while linked to one another through market-clearing conditions. We show the existence and uniqueness of the equilibrium solution. We propose a modified projection method for solving the Nash equilibrium as a variational inequality. We report the numerical results for a case study involving the EVB manufacturing system in Kentucky. Future work will extend the current model to study various trade, government incentives and environmental policies.

Monday Poster

Building A Level 1 Exhibit Hall A3

Monday Poster Session

Poster Session

1 - Project Portfolio Selection at a Brazilian Investment Bank: a Multicriteria Decision-Making Approach Using the Promethee V Method

Andréia Kaori Io Imai Gomes, Aeronautics Institute of Technology, São José dos Campos, Brazil, Arsênio Gomes Neto, Jonas Almeida

The selection of project portfolios in contexts with limited availability of specialists is a recurring and challenging activity within organizations. In this context, this study presents a multicriteria model to support the selection of project portfolios in a Brazilian investment bank. Accordingly, the modeling was conducted using a structured approach consisting of 12 steps, encompassing all stages from stakeholder definition to the final decision recommendation. To this end, the PROMETHEE V method was chosen for being aligned with the non-compensatory rationality of the decision-maker and for allowing explicit consideration of constraints, in this case, the maximum number of specialists available to execute the projects. Following this, seven criteria were defined for the application of the model: "Criticality," "Strategic Area," "Time in Backlog," "Estimated Duration," "Technical Capacity," "Operational Efficiency," and "Priority." In this context, the consequence matrix was composed of 10 alternatives, corresponding to the projects eligible for the portfolio. As a result, an optimal portfolio (P8) was identified, consisting of eight projects and maximizing the adapted net outranking flow. The robustness of the solution was assessed through a sensitivity analysis with $\pm 10\%$ variations in the criteria weights and in the values for the "Estimated Duration" criterion. Based on these results, we observed that portfolio P8 was reproduced in 99.70% of the simulated scenarios, with no occurrence of infeasible cases, which demonstrates the high stability of the solution.

2 - A Multimodal Retrieval-Based Framework for Product Popularity Prediction in Fast Fashion

Sophie Liu, The University of British Columbia, Vancouver, BC, Canada, Xinhui Yu, Chunhua Wu

Accurate product popularity prediction is critical for reliable demand forecasting and inventory management, particularly in the fast fashion clothing sector. Existing approaches mainly employ computer vision techniques to extract high-dimensional visual features from clean clothing product images for predicting popularity, and often overlook other influential factors. As a result, they are prone to overfit and offer limited explainability. To address these challenges, we propose a multimodal retrieval-based framework for predicting product popularity that leverages design and pricing features. Specific fashion-related questions are used as prompts for ChatGPT to generate structured textual descriptions for each product. These descriptions, along with the corresponding product images, serve as inputs to a multimodal retrieval model. The retrieval module returns the top-k most similar popular and unpopular items for a given query, and we use similarity scores as features instead of raw image embedding to reduce input dimensionality. In addition to these similarity scores, we incorporate numerical features (e.g., discount level) and categorical features to build a multi-factor popularity prediction model. Through a case study on SHEIN clothes, our experimental results demonstrate the effectiveness of our proposed framework, highlighting its potential for supporting sustainable production strategies, and indicating the importance of multiple factors in shaping product popularity. Our findings support the employment of AI-driven decision-making to minimize overproduction and overstock in the fashion industry, with the potential to apply to broader fashion categories.

3 - A Robust Multi-Objective Model for Sustainable Closed-Loop Supply Chain of Home Appliances Under Demand and Return Uncertainty

Jorge Restrepo Diaz, Toronto Metropolitan University, Toronto, ON, Canada, Saman Amin

Corporate Social Responsibility (CSR) initiatives are increasingly influencing consumer behavior and shaping the management of Closed-Loop Supply Chains (CLSCs), the systems that integrate forward and reverse logistics in response to business, governmental, and societal pressures. As sustainability expectations rise, organizations are striving to incorporate environmental and social considerations into CLSC design. However, these efforts face challenges due to the often qualitative and uncertain nature of such metrics. This research addresses these challenges by proposing a comprehensive approach for designing a sustainable CLSC for home appliances, products essential for sustainable consumption that are impacted by rapid technological change and shortened product life cycles. The proposed model includes a Regime-Switching Lattice (RSL) framework to account for demand uncertainty and employs Material Flow Analysis (MFA) to assess return rates and the conditions of used products. Furthermore, a novel Multi-Objective Mixed-Integer Linear Programming (MOMILP) model is developed and integrated to simultaneously optimize economic performance, environmental impact, and social well-being. To solve the proposed model and generate Pareto-optimal solutions, this study utilizes an augmented ϵ -constraint method along with dynamic programming. The model is validated using real-world data from the three Canadian provinces of British Columbia, Ontario, and Quebec, and provides a robust decision-support framework for policymakers and supply chain managers aiming to design adaptive and sustainable CLSC for durable goods. The results show managerial insights into product allocation, facility location decisions, and reverse logistics strategies.

4 - Platform Competition and Incentives for Content Creators

Ruomeng Hu, Tianjin University, Tianjin, China, People's Republic of, Dan Ma, Nan Feng

Content platforms traditionally only offer monetary rewards to high-value creators who capture greater time engagement from content consumers. Recently, however, some platforms have begun extending their rewards to low-value creators. This study develops an analytical model to examine how such reward extensions affect creators' participation incentives, end-users' content consumption behavior, and platforms' profitability in a duopoly competition market. We consider two competing platforms that decide whether to extend rewards. Creators choose which platform(s) to join, while end-users choose a platform and allocate their time across creators. We find that reward extension might change creators' platform choices. Interestingly, the ultimate benefits may go to either the focal platform (due to competitive advantage) or/and the rival platform (due to the free-riding effect), depending on the degree of platform differentiation as well as the creators' costs of content production and multi-homing. In addition, we show that extending rewards to low-value creators can even benefit high-value creators when the platforms are highly differentiated, content is less substitutable, and the value gap between creators is small. When analyzing platforms' reward-extension decisions, we find that neither/one/both platforms extending rewards may all turn out to be possible equilibrium outcomes. For example, when low-value creators contribute reasonably significant revenue and when both content production costs and multi-homing costs are not high, only one platform extending rewards could be the equilibrium. Notably, platforms might face a prisoner's dilemma: under certain conditions, both will choose to reward low-value creators, even though both would be better off if neither does.

5 - Evaluation Methodology for Transfer Hub Capacity in Multi-Mode Rail Transit Systems Oriented to Coordinated Operation

Qinghui Xu, Tongji University, Shanghai, China, People's Republic of

As the core nodes connecting different types of rail transit systems, transfer hubs play a crucial role in determining the efficiency and service quality of multi-modal coordinated operation. However, current methods for evaluating the capacity of transfer hubs under multi-system conditions remain relatively coarse, lacking systematic, dynamic, and operation-oriented adaptability. This study, grounded in the concept of coordinated operation, develops a comprehensive capacity evaluation methodology tailored to transfer hubs within multi-mode rail transit systems. By integrating transfer behavior modeling, passenger flow-facility coupling analysis, and the correlation modeling of operational parameters, the proposed approach considers both static facility configuration and dynamic operational scheduling. Using empirical case studies from typical metropolitan transit hubs, the method is validated in terms of its effectiveness in improving capacity estimation accuracy, supporting facility optimization, and informing operational adjustments. The research provides theoretical support and technical tools for the coordinated optimization of regional rail transit networks and for investment decision-making regarding transfer hub infrastructure.

6 - Information Acquisition, Portfolio Choice, and Diversification

Jiming PAN, City University of Hong Kong (CityUHK), Shenzhen, China, People's Republic of, Xiao Qiao, Moris Strub

We study the joint decision of information acquisition and portfolio choice for a mean-variance investor. The investor is concerned with risk from both the information acquisition and investment periods. Optimal policies indicate that the investor should learn about assets with lower prior squared Sharpe ratios in order to improve the prospects of a diversified portfolio, because a concentrated portfolio is associated with an increased risk of extreme returns. Our results show that diversified portfolios can arise in a model without perfect information, suggesting that diversification has wider applicability than previously recognized.

7 - Can Algorithmic Opacity Tame the Generative AI Pricing and Leviathan? Endogenous Consumer Adaptation in AI-Driven Markets

li su, shanghai jiaotong university, shanghai, China, People's Republic of

Generative AI-driven dynamic pricing is transforming corporate pricing strategies, with algorithmic transparency critically shaping consumer adjustments in strategic waiting behavior under real-time AI-optimized pricing. This study develops a two-stage game-theoretic framework that integrates generative AI pricing algorithms with an endogenous time preference model to analyze strategic interactions between consumers and AI systems. We find that (i) While reinforcement learning-based pricing induces rational consumer expectations, algorithmic interpretability asymmetrically affects strategic behavior costs. Moderate opacity in pricing rules elevates prediction complexity, steering consumers toward myopic purchasing decisions and achieving mutual benefits across corporate profits, consumer surplus, and social welfare. (ii) Unlike static pricing models, AI-driven dynamic pricing mitigates price commitment inefficiencies through adaptive feedback mechanisms. However, excessive transparency reduces strategic deliberation costs, reinstating suboptimal "prisoner dilemma" equilibrium. (iii) Strategic costs can be endogenously amplified through complex pricing patterns (e.g. non-monotonic discounts), which systematically deter high-valuation consumers from strategic waiting. The results advance the design of AI-compatible pricing architectures by demonstrating how controlled information asymmetry optimizes market outcomes. Furthermore, the analysis reveals the dual role of generative AI as both an efficiency enhancer and a potential market destabilizer, contingent on algorithmic transparency levels. These insights inform regulatory frameworks for AI pricing systems, emphasizing context-dependent standards to balance innovation incentives with consumer protection imperatives.

8 - Robotic Slot-Shuffling Parking Reservation System: a Demonstration-Augmented Gflownet Approach for Revenue Optimization

hanyu qiu, chinese university of hong kong shenzhen, Shenzhen, China, People's Republic of, Pengyu Yan, Hongke Xie

Abstract: This study introduces a novel parking-reservation system supported by slot-shuffling robots. The system enables drivers to reserve parking slots in advance within a limited planning horizon, divided into multiple time periods. During each period, the system processes at most one reservation request with randomly parking duration (type) and decides whether to accept or reject it to maximize total revenue over the horizon. We model this problem as a Markov Decision Process incorporating a

robotic slot-shuffling mechanism to enhance resource utilization. To address demand uncertainty, we propose a model-free framework called Demonstration-Augmented Replay GFlowNet (DAR-GFN). This framework leverages offline first-come-first-served (FCFS) demonstrations to establish a robust navigation prior for the learning agent, while a prioritized replay buffer of top-revenue online trajectories is employed to update the joint backward policy, thereby improving sample efficiency and resilience to stochastic reservation arrivals. Numerical experiments using real-world parking demand data demonstrate that DAR-GFN significantly outperforms both the FCFS rule and deep reinforcement learning baselines from the literature.

9 - Topologically Enhanced Surrogate-Assisted Genetic Programming for Dynamic Flexible Flow Shop Scheduling with Parallel Machines

Yu Sun, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, China, People's Republic of, Ying Meng

Dynamic Flexible Flow Shop Scheduling with Parallel Machines (DFFSPM) is a challenging NP-hard combinatorial optimization problem critical for efficient resource utilization and productivity in modern manufacturing. Genetic Programming (GP) effectively evolves high-quality dispatching rules for such complex systems but suffers from computationally intensive fitness evaluations. Surrogate-Assisted Evolutionary Algorithms (SAEAs) mitigate this by approximating expensive fitness functions; however, their success largely depends on surrogate model accuracy, significantly influenced by the chosen feature representation. Traditional scheduling features, such as phenotypic characterization, often fail to sufficiently capture the complex characteristics of dispatching rules applied to DFFSPM problems. To overcome this limitation, this study introduces novel topological features derived through Topological Data Analysis (TDA) to enhance surrogate model accuracy. Specifically, during the GP process, candidate dispatching rules are evaluated using small-scale scheduling instances to efficiently generate representative resulting schedules. TDA techniques are then applied to extract robust topological features from these schedules, capturing deeper structural properties of the scheduling solutions. These extracted features form the basis for training more accurate surrogate models, significantly improving predictive capabilities. Numerical experiments on diverse DFFSPM problems with varying objective functions demonstrate that surrogate models trained with topological features achieve superior prediction accuracy. Furthermore, the enhanced surrogate-guided genetic programming framework consistently discovers higher-quality dispatching rules more efficiently, underscoring the substantial potential of integrating TDA within surrogate-assisted evolutionary algorithms.

10 - A Three-Phase Method for Systematically Computing Multiple Optimal Solutions of Nonlinear Constrained Optimization Problem

Cuiqing Zhang, Sun Yat-sen University, Guangzhou, China, People's Republic of, Yusi Zhang

In this article, a novel Three-phase method is proposed to systematically compute multiple local optimal solutions (LOSs) for constrained nonlinear optimization problems. It locates the feasible connected component (FCC) by reformulating the problem as the identification of the stable equilibrium manifold (SEM) of a Quotient Gradient System (QGS) and transforms the search for a LOS into the identification of stable equilibrium points (SEP) of a Projected Gradient System (PGS). In phase I, a FCC is identified via the QGS and a LOS is obtained on this component using the PGS. Phase II enables transition from the current FCC to another one. In Phase III, the PGS is again applied to find additional LOSs simultaneously on the new FCCs. This method offers systematic and deterministic exploration of multiple solutions, does not rely on initial values, and exhibits high computational efficiency. A theoretical foundation for both phase I and III is developed, and these theoretical results are quite general on their own. The proposed method is numerically evaluated to compute multiple LOSs, for instance, a step-by-step demonstration using a 2-D example. In addition, compared with interior point method, heuristic method, and other multiple-solution techniques, the proposed method can efficiently find the multiple or even global optimal solutions across various optimal power system problems with dimensions ranging from 10 to 1000.

11 - Dynamic Causal Modeling and Analysis of Risk Evolution in Vehicle-Nonmotorized Vehicle Interactions for Enhanced Driving Safety

zhangcun Yan, Tongji University, Shanghai, China, People's Republic of

In complex traffic scenarios characterized by high-density mixed flows of motorized and non-motorized vehicles, significant safety hazards exist that not only endanger vulnerable road users but also pose serious challenges for the practical implementation of autonomous driving technologies. To address this critical issue, this study innovatively develops a dynamic causal modeling and analysis framework designed to systematically investigate the generation mechanisms of risk sources and their causal pathways in the evolution of microscopic driving behavior risks during vehicle-non-motorized vehicle interactions. Specifically, this research makes three primary innovative contributions: First, we pioneer an eight-dimensional risk perception model for shared spaces that quantitatively characterizes the dynamic temporal features of vehicle-environment interactions through multidimensional feature parameters. Second, we introduce the advanced PCMC1 algorithm for dynamic causal inference, constructing both Anticipated Collision Time (ACT) and Conflict Risk Field (CRF) models based on interaction data from unprotected left-turn scenarios. Third, through time-lag effect analysis and causal pathway exploration, we reveal the intrinsic causal mechanisms between feature variables and conflict risks. Empirical results demonstrate that the dynamic CRF model exhibits superior risk evolution characterization capability, clearly identifying both positive reinforcement effects and negative attenuation effects in risk transmission. Notably, the dynamic causal CRF model achieves 1-2 second advanced risk prediction and precise risk source localization, providing reliable decision-making support for warning systems. These findings not only establish an innovative theoretical framework for driving safety warnings in complex mixed traffic environments, but also lay a crucial technical foundation for optimizing intelligent driver assistance systems.

12 - System Modeling of the U.S. Aluminum Supply Chain: Integrating Dynamic Material Flow Analysis and Life Cycle Assessment

Sidi Deng, University of Michigan, Ann Arbor, MI, United States, Daniel Cooper

The aluminum industry is a cornerstone of the global economy and contributes substantially to greenhouse gas (GHG) emissions. The ever-growing demand for aluminum underscores the need for a sustainable and low-carbon supply chain. Recognizing that regional contexts play a critical role in shaping supply chain dynamics, this study narrows its scope to U.S. aluminum production and consumption. An object-oriented model is established as a digital representation of the U.S. aluminum supply chain, which serves as a computational engine for simulating dynamic material flows and quantifying time-dependent carbon footprints associated with both domestic production and international trade. This presentation intends to showcase the emission performance outcomes from the model under various scenarios, featuring temporal emission breakdowns by supply chain stage, carbon source, and commodity, as well as the cradle-to-gate carbon footprint of each end-use product category. Three decarbonization pathways (e.g., electrification, hydrogen, and inert anodes) have been encoded into the model, with their performance evaluated from 2020 to 2050 under evolving energy profiles and climate-change initiatives. This study underscores the critical role of integrating material flow analysis and life cycle assessment in guiding sustainable supply chain planning and industrial decarbonization. It introduces a system-oriented analytical paradigm applicable to other material and energy intensive industries.

13 - A Two-Echelon Energy-Aware Routing Model for Truck–Drone Coordination in Mountainous Fresh Logistics

shunfeng zhang, Southeast University, China, Nanjing, China, People's Republic of, Xinlian Yu, xiaofeng li, haijun mao

The logistics of fresh agricultural products in mountainous regions are confronted with significant challenges, including complex terrain and weather conditions, limited cold chain and charging infrastructure, and low delivery efficiency. To address these challenges, we propose a two-echelon energy-aware truck-drone routing problem (TEDR). Specifically, trucks are responsible for direct deliveries to accessible demand points and serve as mobile bases for drone operations. Drones are dispatched from truck stops to reach multiple demand points that are inaccessible to trucks due to terrain or road constraints. A bi-level decision framework is developed to ensure effective coordination between trucks and drones, considering dynamic energy consumption. The upper level employs a CNN-Transformer hybrid model to predict drone energy consumption in real time, accounting for variations in terrain and weather conditions. The lower level constructs a bi-objective optimization model to minimize the total operational cost and the maximum delivery time, thereby enabling adaptive path planning for the truck-drone system.

To solve this complex problem, we design a Deep Reinforcement Learning-enhanced Multi-Objective Evolutionary Algorithm (DRL-MOEA). A case study using real-world data from the mountainous region of Kunming, China, demonstrates the effectiveness of the proposed approach. Our approach significantly reduces cost and time while improving solution quality and Pareto front coverage over conventional methods. System performance regarding key parameters such as energy consumption rate, delivery reliability, and operational robustness is also analyzed. These results confirm the effectiveness of the proposed framework and offer valuable insights into intelligent cold chain logistics in complex terrains.

14 - Trajectory Prediction and Risk Assessment in Car-Following Scenarios Using a Noise-Enhanced Generative Adversarial Network

Xinyi Li, The Hong Kong Polytechnic University, Hong Kong, China, People's Republic of, Ting Fu

Traditional conflict analysis methods, relying on the assumption of constant velocity, often fall short in capturing the dynamic nature of driver behavior randomness during the interaction process. Predicting all potential collision trajectories proves crucial for comprehensive safety analysis. To address the challenge of accounting for trajectory randomness in car-following scenarios, this study introduces a noise-enhanced generative adversarial network, named Car-Following GAN, designed for predicting collision trajectories based on data from the Shanghai Naturalistic Driving Study (SH-NDS). The model employs an encoder-decoder framework, integrating a noise enhancement module to capture the intrinsic randomness of driving patterns. Demonstrating notable robustness across varying environmental conditions, our model showcases adaptability for trajectory prediction in diverse driving scenarios. A conflict measure, termed the Rear-end Collision Risk Index based on Car-Following GAN (RCRIC), is proposed to quantify the risk of a rear-end collision. Our approach conducts a comprehensive case analysis to assess the impact of various traffic risk factors on RCRIC. The results underscore that our noise-enhanced approach significantly improves the trajectory prediction accuracy of the model when compared to other noise addition methods. This enhancement is observed across various prediction time windows and under different weather conditions. Moreover, RCRIC, derived from the model employing our noise-enhanced approach, effectively mirrors the dynamics of rear-end collision risk by explicitly incorporating trajectory randomness into its assessment. Furthermore, the findings underscore the significant influence of light conditions, traffic density, and weather conditions on driving risk.

15 - How Does China's Iron and Steel Industry Achieve Carbon Emission Reduction Goals? from the Perspective of Production Processes Optimization

Chen Wang, University of Science and Technology Beijing, Beijing, China, People's Republic of, Shufen Dai, Wei Gu

With the advancement of global carbon neutrality goals, the steel industry faces increasing pressure to reduce emissions. The limited scrap resources and high steel production demand necessitate the long-term coexistence and gradual substitution between the Blast Furnace-Basic Oxygen Furnace (BF-BOF) process and the Electric Arc Furnace (EAF) process in global steel industry. Therefore, a scientific evaluation of the emission reduction potential of BF-BOF and EAF process substitution and optimization, and the development of emission reduction pathways are of great significance for achieving the "dual carbon" goals.

Taking Chinese steel industry as an example, this paper explores the low-carbon technology pathway planning from 2023 to 2050 using a machine learning method and scenario analysis from the perspective of production processes optimization. We found that process optimization has played a role in reducing carbon emissions in the steel industry over the past decades. And by the low-carbon potential prediction of technologies, we found that the BF-BOF-based technologies have limited effectiveness, with carbon emissions decreasing from 2,199.95 MtCO₂ in 2023 to 1,025.27 MtCO₂ in 2050; while the EAF process scenario shows significant promise. The co-abatement scenario, supported by various technologies, represents a reasonable emission reduction pathway, projected to be more efficient and effective than that in the EAF process scenario. However, despite this, low-carbon technologies for process optimization are still unable to achieve the 2060 carbon neutrality target. Therefore, this paper also proposes the application strategies of mainstream emission reduction technologies for different processes at different stages while providing specific policy recommendations.

16 - Improving Front Distribution Center Fulfillment Rates: A Distributionally Robust Approach

Qingwei Jin, Zhejiang University, Hangzhou, China, People's Republic of, Haodong Feng, Man Feng, Qianqian Jin, Xinru Hao, Yidong Zhang, Lei Cao

In E-commerce distribution networks, front distribution centers (FDCs) are extensively employed to reduce delivery time, which has a significant impact on customers' purchase intentions and loyalty. Upon customer order placement, the corresponding FDC promptly fulfills the order, ensuring a short delivery time. If there is a shortage in the FDC, the order is then fulfilled by the regional distribution center (RDC) with a longer delivery time. Otherwise, a lost sale occurs. A key performance metric is the FDC fulfillment rate which reflects the proportion of orders successfully fulfilled by FDCs. In this paper, we design a distributionally robust allocation model that improves FDCs' fulfillment rates and, at the same time, maintains the region's overall fulfillment rate. We transform this model into an equivalent mixed integer second order conic programming (MISOCP) model, and an approximate mixed integer linear programming (MILP) model by partitioning the robust domain. Through numerical experiments, we investigate the impact of the balance coefficient on fulfillment rates and demonstrate the excellent performance of our model in a rolling horizon setting, particularly when faced with inaccurate demand forecasts. We implement our model within the distribution network of the home appliance industry of Tmall platform (the largest E-commerce retail platform in China), resulting in a notable improvement in the FDC fulfillment rate (exceeding 10%) and a substantial boost in gross merchandise volume (GMV).

17 - Research on Graph Convolutional Recommendation Algorithm Based on Hyperbolic Space and Contrastive Learning

Pengyu Lu, Harbin Institute of Technology, Harbin, China, People's Republic of, Zicheng Wang

To address the limitations of existing recommender systems in capturing hierarchical user preferences and handling long-tail item recommendations, we propose Hyperbolic Contrastive Graph Convolutional Networks (HC-GCNs). This architecture, built upon a 3-layer LightGCN and HyperGCN baseline, leverages hyperbolic space embeddings and contrastive learning to mitigate oversmoothing and data sparsity in graph-based recommendation models. We employ the Lorentz model to embed user representations in hyperbolic space, thereby capturing hierarchical relationships and enhancing long-tail item recommendation. Triplet contrastive learning is implemented to refine embeddings by minimizing the distance between user-target item embeddings and maximizing the distance between user-non-target item embeddings. We evaluate HC-GCNs using AUC, F1@K, Recall@K, and NDCG@K metrics on the MovieLens-10M and Amazon-Book datasets. Experimental results demonstrate that HC-GCNs outperform the baselines across all evaluation metrics. Specifically, on the Amazon-Book dataset, HC-HyperGCN exhibits a performance improvement exceeding 12.8% compared to HyperGCN.

18 - A Collaborative Production Planning Game in Steel Industry

Yuming Zhao, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, China, Shenyang, China, People's Republic of, Shengnan Zhao

In this work we introduce a collaborative production planning game inspired by real-world joint production processes in the steel industry, which requires complex coordination among companies of the same production stage but with varying capacities and demands, aiming to improve production efficiency and reduce costs. We introduce a column-generation-based algorithm specifically designed to identify a core allocation of among cooperating participants. A heuristic is proposed to solve the planning subproblem, which considers grouping, carryover, and backlogging in production processes. Numerical experiments are carried out and demonstrate the effectiveness of proposed models. The results indicate that the collaborative framework makes improvements in resource utilization and ensures fair and stable cost-sharing among participants.

19 - How Digital Transformation Shapes Firm Performance: a Meta-Analytic Integration of Resource-Based and Institutional Perspectives

Hongyu DUAN, Hong Kong Baptist University, Hong Kong, Hong Kong, Chenming Peng, Danny Wang, Yulin Fang

Digital transformation (DT) has become a critical driver of firm performance, enabling organizations to enhance operational efficiency, foster innovation, and strengthen customer engagement. However, despite its potential, the performance outcomes of DT vary significantly across contexts, highlighting the need to better understand the factors that shape this relationship. Drawing on the Resource-Based View (RBV) and institutional theory, this study examines how the configuration of DT strategies, defined by technology type (incremental vs. disruptive), scope (organization-wide vs. department-wide), and strategic orientation (impact-driven vs. effort-driven), moderates the DT-performance relationship. Additionally, we explore how institutional factors, including education level, internet penetration, power distance, rule of law, and economic growth, condition the effectiveness of DT. Using a meta-analytic approach, we synthesize 783 effect size estimates from 225 empirical studies conducted across 36 countries. This study contributes to the literature by integrating strategic and contextual perspectives to explain heterogeneity in DT outcomes. Practically, the findings offer evidence-based guidance on how firms can strategically configure digital initiatives and align them with institutional environments to enhance firm performance.

20 - Network-Based Pick-up and Drop-off Optimization for Shared Autonomous Vehicles

Sunyu Wang, Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, Hong Kong, Fangyi Ding, Yineng Wang, Anthony Gar-On Yeh, Guan Huang

Shared autonomous vehicles (SAVs) are reshaping the mobility system, with ride-hailing services being regarded as a pioneering application market. Despite its benefits, SAVs inevitably exacerbate the competition for limited curb space, which is already a scarce resource. Designated pick-up and drop-off sites, if reasonably planned, present a promising solution to this challenge. In this paper, we propose a novel network-based data-driven approach to optimize PUDO locations for SAVs. Network kernel density estimation (NKDE) is used to measure the PUDO demand along the road network based on historical ride-hailing data. The results are fed into a maximal coverage location problem (MCLP) model to optimize the spatial distribution of PUDO sites by maximizing the demand they can cover. The spatial optimization model is used both prescriptively and descriptively, assessing transport policies and enhancing the operations of SAVs. Results show that our model outperforms the baseline models. We also reveal that a suitable PUDO site density is about 20/10 km², with denser sites in bustling urban centre and sparser sites in suburbs. The framework is highly generalizable and reproducible and can be extended to future studies.

21 - Asymptotic Optimality of Simple Policies for Serial Inventory Systems with Lost Sales: Beyond Echelon Base-Stock Policy

Huanyu Yin, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Xiting Gong

We consider a periodic-review serial inventory system, where demands are satisfied by the on-hand inventory of the lowest stage, and unsatisfied demand is fully lost. The objective is to minimize the long-run average holding and penalty costs. When unsatisfied demand is backlogged, [Clark and Scarf \(1960\)](#) prove that the Echelon Base-Stock (EBS) policy, maintaining a constant echelon inventory position in each period, is optimal. However, for the lost-sales setting, the optimal replenishment policy is notoriously complex and intractable due to the curse of dimensionality. Previous study focuses on the asymptotic analysis of the best EBS policy, which performs poorly when lead times are large or the unit penalty cost is relatively small. Hence, we propose two new classes of heuristic policies: Echelon Projected Non-stockout Probability (EPP)₃ and Echelon Project Inventory Level (EPIL) Policies. We prove that the best single-index EPP ₃ and EPIL policies are asymptotically optimal with large unit penalty costs under a broad class of demand distributions. Moreover, as external lead times increase, we establish a constant asymptotic bound on the relative optimality gap of the best dual-index EPIL policy, which admits a closed-form expression under exponential demand distributions. We also analyze how this gap behaves under strong demand pooling. To facilitate computation, we develop a sequential optimization method to obtain the sequentially optimized EPP ^{SO}₃ policy. Our numerical results show that all policies perform close to optimal and significantly better than the EBS policy.

22 - TCLMA: A Two-Dimension Contrastive Learning Based Multiagent Framework for Invention Novelty Evaluation

Lin ZHENG, City University of Hong Kong, Hong Kong, Hong Kong, Qingyu Xu, Jian MA

AI-supported methods are increasingly employed to measure human knowledge outputs, such as inventions. However, evaluating the creativity of inventions remains a persistent challenge. The notion of creativity, particularly invention novelty, is crucial for distinguishing significant scientific and technological advancements, spanning activities from basic research to groundbreaking innovations. Recently, the emergence of Large Language Models (LLMs) has been explored into their potential for assessing invention novelty. However, existing models depend heavily on implicit reasoning, which causes them to lean on the experiential knowledge encoded in LLMs. Meanwhile, time is a critical factor in determining whether an invention is truly new, a capability that current LLM-based methods inherently lack. This paper introduces TCL-MFSNE, an advanced multi-agent framework designed to address these challenges in invention novelty evaluation for patent assessments. TCL-MFSNE incorporates a two-dimensional contrastive learning module—comprising positive-negative contrastive learning for enhanced discrimination and temporal contrastive learning to account for the time-sensitive nature of invention novelty. By integrating these modules with explicit Chain of Thought (CoT) reasoning, TCL-MFSNE enables LLMs to systematically evaluate invention novelty while grounding their reasoning in both contemporary and historical contexts. Furthermore, the framework employs Byzantine Consensus theory, ensuring robust multi-agent collaboration for stable and accurate decision-making even in the presence of inconsistencies. Experimental results demonstrate that TCL-MFSNE achieves state-of-the-art performance, with an accuracy and F1-score of 85% on patent classification tasks. By addressing real-world challenges in invention novelty assessment, TCL-MFSNE advances the theoretical capabilities of LLMs and provides a practical tool for improving intellectual property evaluations.

23 - Operation Strategy for Styling Box Platform

Yuan Qu, University of Hong Kong, Hong Kong, Hong Kong, Max Z.J. Shen

Styling box selling has emerged as a novel online shopping format, shifting from traditional active browsing to a passive model where customers periodically receive curated selections of items. Customers can keep desired items and return the rest, often benefiting from a discount if they retain the entire box. This study investigates the platform's operational strategies—discount, recommendation, and pricing—critical to this unique business model. We develop a framework under the Stackelberg setting that captures the interplay between a profit-maximizing platform and strategic buyers with heterogeneous preferences. First, we analyze the buyer choice model and derive the corresponding platform profit model. Our analysis reveals some dominant item swap properties, which enable the discussion of the optimal recommendation strategy and characterize the properties of the optimal discount policy. Under this recommendation strategy, we further examine the optimal discount policy in a homogeneous pricing setting, providing a stylized model that offers a macro-level perspective and a simple estimation method for platform decision-making. Finally, numerical experiments validate our theoretical findings and explore the implications for optimal pricing strategies.

24 - Continuous and Monotone Bayesian Nash Equilibrium with Incomplete Information About Player's Risk Preferences

Ziheng Su, The Chinese University of Hong Kong (CUHK) - Department of Systems Engineering & Engineering Management, Hong Kong, Hong Kong, Huifu Xu

In this paper, we consider a non-collaborative game where each player faces two types of uncertainties: aleatoric uncertainty arising from inherent randomness of underlying data in its own decision-making problem and epistemic uncertainty arising from lack of knowledge and statistical information on the rivals' risk preferences. By assuming that players are risk-averse against aleatoric uncertainty and risk-neutral on the epistemic uncertainty, we propose a Bayesian risk minimization model to describe players' interactions. We investigate existence and uniqueness of a continuous monotone equilibrium arising from the game, termed CMBNRE, where "R" is used to emphasize the game is concerned with risk minimization as opposed to utility maximization in the existing BNE models. We derive sufficient and/or necessary conditions for ensuring existence of CMBNRE from two perspectives: monotonicity of a continuous BNRE and continuity of a monotone BNRE. Continuous BNE emphasizes continuity of each player's response function with respect to variation of the type parameter whereas monotone BNE focuses on monotonicity of the player's response function and/or rival's response functions. The former is derived by virtue of Schauder's fixed point theorem and the latter is established by other fixed point theorems based on single-crossing and quasi-supermodularity, the proposed CMBNRE model synthesizes the two modeling frameworks. We discuss numerical methods for computing an approximate CMBNRE from stochastic optimization perspective and apply the proposed model and computational schemes to a reinsurance competition problem. The test results provide some insights about how indemnity parameter may affect reinsurer's behaviour, market share and profitability.

25 - Condition Number Shrinkage by Joint Distributionally Robust Covariance-Precision Estimation

RENJIE CHEN, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Huifu Xu, Viet Anh Nguyen

We propose a distributionally robust formulation for the simultaneous estimation of the covariance and precision matrix of a random vector. The proposed model minimizes the worst-case weighted sum of Stein's loss of the precision matrix estimator and the Frobenius loss of the covariance estimator against all distributions from an ambiguity set centered at the empirical distribution. The radius of the ambiguity set is measured via convex spectral divergences. We show that the proposed distributionally robust estimation model reduces to convex optimization problems and thus gives rise to computationally tractable estimators. The estimators, as an outcome of distributional robustness, are shown to be nonlinear shrinkage estimators. The eigenvalues of the estimators are shrunk nonlinearly towards a scalar matrix, where the scalar is determined by the weight coefficient of the loss terms. We show that the shrinkage effect improves the condition number of the estimator. We provide the explicit expression of the shrinkage estimators induced by Kullback-Leibler divergence, Wasserstein divergence, Symmetrized Stein divergence, and Frobenius divergence. We also provide a parameter-tuning scheme that adjusts the shrinkage target and intensity that is asymptotically optimal. Numerical experiments on synthetic and real data show that our shrinkage estimators perform competitively against state-of-the-art estimators in practical applications.

26 - Navigating Endogenous Uncertainty in the New Energy Paradigm: Renewable Integration and Resilience Enhancement

Yunhe Hou, The University of Hong Kong, Hong Kong, Hong Kong

The accelerating integration of renewables and the rising frequency of extreme events have fundamentally shifted the uncertainty landscape in power systems. Classical reliability-oriented frameworks, designed for stable conditions, increasingly struggle as uncertainties from renewable variability and extreme events account for a dominant share of system risks. We will introduce resilience-oriented scheduling and planning approaches that break traditional reliability constraints by embracing and proactively managing these high-impact uncertainties. A central focus is on identifying and modelling Decision-Dependent Uncertainty (DDU), where strategic decisions—such as renewable siting, grid reinforcement, or maintenance scheduling—directly shape future uncertainties by altering congestion patterns or failure probabilities. To address the dimensionality and non-convexity challenges inherent in DDU, we present new computational methods based on scenario-based reformulations and tractable approximations. Case studies demonstrate how these approaches enable power system planners to anticipate, influence, and mitigate extreme-event-driven uncertainties, offering a transformative path toward resilient, renewable-rich grids.

27 - Price Competition in a Duopoly with Three Vertically Differentiated Products

Masashi Umezawa, Tokyo University of Science, Chiyoda-ku, Tokyo, Japan, Ryoga Ishihara

This study investigates price competition in a duopoly market where each firm offers three vertically differentiated products—low, medium, and high quality. While existing research has primarily focused on product lines with only two types, considering three quality levels is more realistic, given the presence of a middle segment of consumers. The paper addresses several key questions: How do firms set prices for three products while mitigating cannibalization? How do consumer preferences and production costs affect equilibrium prices and profits? Do equilibrium profits rise when firms offer three products instead of two? If so, under what conditions?

The paper develops a model of horizontal and vertical differentiation in a duopoly and yields two primary findings. First, regarding comparative statics, while higher marginal costs typically reduce profits, the analysis shows that, under certain conditions, increasing the marginal cost of one product can increase profits due to strategic interdependence and demand substitution across products. Specifically, when the marginal cost of the medium-quality product (M) is low, profits can rise with increases in the marginal cost of either low (L) or high (H) quality goods. When M's cost is high, increasing M's cost can also raise profits.

Second, comparing two-product and three-product strategies, the study identifies conditions under which firms earn higher profits with three products. These depend on the relative differences in marginal costs across goods and on which product becomes more cost-effective to sell. The analysis provides clear thresholds for these outcomes.

28 - Structured I/O Design for Modular Scheduling Experiments

Juntaek Hong, Pohang University of Science and Technology, Pohang, Korea, Republic of, Kangbok Lee

This study proposes a structured I/O approach for scheduling experiments, aiming for rapid development and traceability. The modular use of multiple DataFrames reflects the inherently tabular structure of widely-used benchmark datasets, reducing the overhead of reading and transforming raw text into model-ready formats. Traceability is enhanced through a solver logger that captures parameter settings and performance metrics over time. This also approach supports modular experimentation and transparent reporting. Future work includes open-sourcing tools and broader validation across commercial solvers.

29 - Developing a Screening Methodology for Dedicated Freight Truck Lanes Using Transportation Big Data and Traffic Simulation

GwanYong OH, Inha University, Incheon, Korea, Republic of, Sungmin Kim, Daejin Kim

This study proposes a methodology for identifying candidate sections for dedicated freight truck lanes by integrating big transportation data with traffic simulation techniques. Utilizing real-world traffic volume data from the Korea Transport Institute's View-T system, the analysis evaluates traffic and environmental factors across the national road network. Thirteen key indicators—including freight vehicle ratio, average speed, congestion intensity, and pollutant emissions—were selected, normalized using Min-Max scaling, and aggregated into a composite Screening Score. Based on this score, high-priority road segments were systematically identified. The prioritization process was further refined by incorporating road attributes such as road classification, with weighted adjustments reflecting freight traffic volume to better capture freight-dense corridors. To assess the potential impacts of implementing dedicated freight lanes, top-ranked sections underwent microscopic traffic simulation analysis using VISSIM, evaluating metrics such as changes in average speed, congestion relief, emissions reduction, and traffic safety improvements. This study presents a practical and data-driven analytical framework that supports both strategic screening and simulation-based policy assessment for freight infrastructure planning.

30 - Deep Learning-Based Staging and Grading of Periodontitis Using Geometric Features from Panoramic Radiographs

Jae Jun Lee, Department of Industrial and data engineering, Hongik University, Seoul, Korea, Republic of, Yong Oh Lee, Kyeong Won Paeng, Jae Kook Cha

Periodontitis is a chronic condition that causes tooth and alveolar bone loss and affects a large portion of the global population. Traditional diagnosis depends on visual assessments by clinicians, resulting in high variability and limited consistency. While deep learning-based approaches have been explored, their diagnostic performance often falls short for clinical use.

We present a deep learning system for automated staging and grading of periodontitis using panoramic dental radiographs. Unlike prior methods that measure periodontal bone loss (PBL) at a single mesial site, our approach computes the PBL-to-CEJ (Cemento-Enamel Junction) ratio at two parallel-offset endpoints of each tooth, improving anatomical accuracy and staging reliability.

Multiple U-Net-based segmentation models were evaluated to extract tooth and PBL/CEJ boundaries, and the best-performing model was used to compute staging scores. The maximum per-patient score determined disease stage, while length-based features enabled longitudinal comparison of follow-up images for clinical grade assignment.

Experimental results demonstrated improved staging accuracy over previous methods and reliable discrimination of severe cases, supporting the system's potential as a clinical decision aid in dental diagnostics and treatment planning.

31 - Optimal Behavior-Based Load Balancing Under Re-Tou Pricing: Understanding Latent Classes’ Temporal EV Charging Preferences

Bo-Eun Choi, Korea university, Seoul, Korea, Republic of

Although renewable energy production as an alternative to fossil fuels is increasing globally, electricity generated from renewable sources is often curtailed because of temporal imbalances between supply and demand in energy systems. In response, demand-side strategies that shift electricity consumption across temporal periods enhance the system efficiency without major supply side alterations or investments. Consequently, strategies focusing on shifting consumer EV charging patterns—transferring demand from nighttime to periods of high renewable energy availability, mainly daytime—are gaining attention. In this context, the heterogeneity of user behavior is considered to identify group characteristics by segmenting consumers. This study employs a discrete choice experiment (DCE) to analyze consumer preferences regarding EV charging behavior, utilizing survey data from approximately 1,033 participants. Subsequently, latent class analysis (LCA) was conducted to identify distinct user segments and their temporal charging preferences.

Specifically, this study proposes Renewable Energy Time-of-Use (RE-ToU) pricing that incorporates real-time renewable energy share information with pricing signals, thereby delivering both economic and environmental signals to consumers. Through temporal charging probabilities according to these signals, we identified group-specific contributions, namely the amount of behavioral shift toward daytime charging periods. Subsequently, an empirical analysis of temporal grid stability is conducted based on the charging demand affected by consumer preferences, enabling the identification of time-varying demand-supply patterns according to demand fluctuations. The findings provide policy implications for designing demand-side management strategies that promote behavioral transitions in renewable energy systems.

32 - The Approximation Ratio of The Multifit Algorithm: a Computer-Assisted Proof

Insoo Park, POSTECH, Pohang, Korea, Republic of, Kyungduk Moon, Kangbok Lee

The identical parallel machine scheduling problem to minimize the makespan is a classic NP-hard problem in scheduling theory. Among its approximation algorithms, the MULTIFIT algorithm, proposed over four decades ago, remains particularly notable. However, its exact approximation ratio remains unresolved for certain numbers of machines, leaving long-standing open questions. In this work, we present a computer-assisted proof framework based on mathematical programming to verify the approximation ratio of MULTIFIT for all previously open cases.

33 - Cost-Effective Planning of Accelerated Degradation Reliability Demonstration Tests Based on Degradation Model Selection

Hyojung Kim, Chosun university, Gwangju, Korea, Republic of, Seong-joon Kim

Reliability Demonstration Tests (RDTs) are widely used during the product development stage to verify whether a product meets its required reliability level. However, for highly reliable products, traditional lifetime-based RDTs often require excessive time and cost, making them impractical for real-world implementation. To address this challenge, Accelerated Degradation Tests (ADTs) have been increasingly adopted, leading to the design of Accelerated Degradation Reliability Demonstration Tests (ADRDTs) that shorten test durations by accelerating the degradation process.

A key element in ADRDT planning is the selection of an appropriate degradation model. This study presents a practical guideline for selecting between two major degradation modeling approaches: the General Path Model (GPM) and the Stochastic Process Model (SPM). Based on the selected model, a cost-effective ADRDT plan is proposed, which explicitly considers both producer's and consumer's risks.

To evaluate the proposed method, sensitivity analyses and operating characteristic (OC) curves are provided using various numerical examples. Additionally, the practical applicability of the approach is demonstrated through a real-world case study using degradation data from an automotive Electronic Control Unit (ECU). The proposed methodology helps reduce testing burden while ensuring statistical confidence in reliability demonstration.

34 - Adaptation of A Multi-Task Learning Strategy for Tool Wear Estimation Paired with Cutting Force

hyein kim, Korea Institute of Industrial Technology, Ansan-si, Korea, Republic of, Soohyun Nam, Jeongin Koo

The deployment of online tool condition monitoring and machine diagnostics is critical to achieving sustainable manufacturing. Accurate tool wear prediction enables automated tool replacement by comparing real-time wear against predefined tool life limits, thus enhancing productivity and reducing costs. This study proposes a multi-task learning (MTL) framework that simultaneously estimates tool wear and cutting force, improving both monitoring accuracy and efficiency. Cutting force, a key indicator of tool condition, machining stability, and surface quality, is monitored in real time. Its mean and variance are used as target variables alongside tool wear, forming a joint prediction objective. Data are collected from multiple sensors and the CNC system, with meaningful features extracted through signal processing techniques. To validate the proposed MTL approach, we develop models based on Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Networks (TCNs). Model performance is evaluated in terms of accuracy, parameter size, and computational cost (FLOPs), and is compared against single-task learning (STL) models. Our results highlight the importance of inter-task correlation, demonstrating significant accuracy gains in the MTL models across architectures. This work provides an effective strategy for applying MTL to machining processes, contributing to intelligent and efficient manufacturing.

35 - Covariance Aligned Gaussian Process

Juhuhn Kim, KAIST, Daejeon, Korea, Republic of, Jong-Seok Lee

Gaussian process (GP) is a widely used statistical modeling technique for its nonparametric flexibility and ability to quantify uncertainty. However, standard GP typically assumes identical training and test distributions, making transfer from abundant source domain to scarce target domain challenging. To address this, we introduce Covariance Aligned GP (CAGP) framework that (1) induces functional priors from source domain, (2) aligns source and target second-order statistics via joint covariance alignment, and (3) performs a Bayesian update on the scarce target observations. By explicitly matching covariance structures, CAGP preserves rich source-domain structure while robustly adapting to the target domain under concept shift. Across multiple benchmarks, CAGP shows competitive predictive performance on scarce target data. This framework can be applied to a wide range of regression problems under concept shift or domain adaptation in simulation and real-world settings.

36 - Research on Optimal Technology Selection Paths for the Global Steel Industry to Reduce Greenhouse Gas Emissions

Yoonmo Koo, Seoul National University, Seoul, Korea, Republic of, Hyunkyung Joo

The global steel industry, responsible for approximately 7% of total greenhouse gas emissions and nearly 30% of industrial sector emissions, is under increasing pressure to decarbonize in alignment with global net-zero targets. With steel demand projected to grow by about 30% by 2050, reducing production is not a feasible mitigation pathway. Instead, the industry must pursue technological shifts such as hydrogen-based direct reduced iron (DRI), electrification, carbon capture, utilization and storage (CCUS), and bioenergy solutions.

This study develops an optimization model using Positive Mathematical Programming (PMP) to quantitatively evaluate the evolution of technology choices in the steel sector across six major producing regions: South Korea, China, Japan, the United States, the EU15, and India. The model incorporates dynamic learning-by-doing effects, allowing technology costs to decline as cumulative deployment increases, thereby capturing the cost competitiveness transition over time. PMP overcomes the corner solution limitations of linear programming by reflecting realistic, multi-technology adoption behaviors in industrial decision-making.

The analysis examines two learning scenarios—presence or absence of learning-by-doing, and differing learning rates based on conservative and progressive assumptions—to assess their impact on low-carbon technology adoption. Learning effects notably enhance the competitiveness of high-cost options like hydrogen and electrification, with adoption timing varying by country due to differences in resources and infrastructure. This research contributes to policy and investment decision-making by offering a dynamic and regionally differentiated perspective on future steel decarbonization strategies. It provides a robust analytical foundation for designing flexible, long-term pathways aligned with carbon neutrality goals.

37 - VPP Bidding and Operation in Electricity Markets Based on Stochastic Dual Dynamic Integer Programming

YoungDoo Choi, Pohang University of Science and Technology, Pohang-si, GYEONGSANGBUK-DO, Korea, Republic of, Sanghyeon Bae, Seokwoo Kim, Dong Gu Choi

The growing penetration of renewable energy in electricity markets has necessitated the introduction of a real-time market between the day-ahead market and actual electricity dispatch. Optimal decision-making for day-ahead and real-time bidding, as well as realtime operations, can be formulated as a scenario-based multi-stage stochastic programming problem. This problem is complicated by the possibility of negative prices due to excessive renewable generation, the non-convex nature of payment functions, and the temporal mismatch between real-time bidding and operations. Given the problem's 25-stage structure and inherent non-convexities, we adopt a formulation amenable to Stochastic Dual Dynamic Integer Programming (SDDiP).

To address the dependency of real-time prices on corresponding day-ahead prices, and to mitigate the high dimensionality of state variables induced by multi-dimensional dayahead prices, we propose a hybrid framework that integrates two-stage stochastic programming with SDDiP. We expect this approach to more effectively capture the coupled structure of day-ahead and real-time markets, and to outperform conventional SDDiP in such integrated market environments.

38 - Assessing Decarbonization Strategies in the Global Cement Industry: A Multi-Regional Optimization Approach

Byoungchan Chun, Seoul National University, Seoul, Korea, Republic of, Jonghyeok Han, Yoonmo Koo

The cement industry accounts for approximately 7% of global industrial greenhouse gas emissions and is considered one of the three major high-emission sectors along with steel and petrochemicals. As such, it is facing increasing pressure to reduce emissions in line with Nationally Determined Contribution (NDC) of each country. This study evaluates the feasibility of emission reduction targets in the cement industries of six major regions: South Korea, China, Japan, the United States, India, and the European Union. Key reduction technologies which are applicable to the cement sector—such as Energy efficiency improvements, Eco-friendly fuels, Decarbonated raw materials, and Carbon Capture, Utilization, and Storage (CCUS)—were identified and analyzed. The assessment was conducted using a process-based optimization model developed in GAMS, considering prediction for each country's demand for cement and emission reduction goals. Based on this framework, the study investigates the emission reduction pathways due to the technology selection and the feasibility of achieving NDC targets. In addition, scenario analyses were performed to assess the impact of policy changes, such as adjustments to emission goals and the introduction of carbon pricing, in order to test the impact of such changes on the emission goals

and pathways. The findings of the study provide a quantitative basis for the insights into the effectiveness of policy instruments such as incentive for some key technologies and introducing carbon taxes.

39 - Structuring AI Service Context: A Framework for Systematic Idea Generation

Geun hyeon Han, Incheon national university, Incheon, Korea, Republic of, Do-Hyeon Ryu

Artificial intelligence-based services (AIS) use autonomous learning and adaptive capabilities to predict user behavior and deliver personalized experiences. However, this innovation complicates the service context, which refers to the circumstances surrounding user-service interactions. As AIS processes various information, including environmental factors and interaction modalities, the number of contextual elements increases dramatically, and their combinations vary across services. Consequently, inadequate consideration of this context during the design phase leads to poor user experience and increases the risk of AIS design failure. To address this challenge, this study proposes a two-step framework for structuring the AIS context and applying it to design. First, GPT-4o-based few-shot information extraction analyzes literature to derive the relevant contextual dimensions and parameters. The extracted information forms a morphological matrix serving as the theoretical foundation. Second, real company cases and news data are mapped to this matrix to construct networks. Network analysis identifies dense connections as established service patterns with potential feasibility, while link prediction algorithms identify potentially viable, unexplored connections as candidate AIS opportunities. This study proposes a framework for structuring AIS context and a data-driven methodology for identifying novel service opportunities to support early-stage design decisions.

40 - A Pod Transfer, Repositioning, and Robot Scheduling Problem in the Robotic Mobile Fulfillment System by Deep Reinforcement Learning

Byung Soo Kim, Incheon National University, Incheon, Korea, Republic of, Yong Jae Kim

A robotic mobile fulfillment system (RMFS) is one of the parts-to-picker systems adopted by many e-commerce companies due to its efficient order processing and flexibility in response to changes in demand. We address a pod transfer, repositioning, and robot scheduling problem in the robotic mobile fulfillment system. The objective is to minimize the makespan, the time the last pod returns to the storage location. To solve the problem, we propose a rule-based deep reinforcement learning algorithm for a dynamic RMFS environment. The reinforcement learning network dynamically selects the decision rules to apply when choosing a storage location for the pod to return to after completing its picking operation. Compared to other rule-based heuristic algorithms and meta-heuristics, the deep reinforcement learning algorithm with several decision rules obtains comparative results in large-sized problems. We analyze the proportion of rules selected in the deep reinforcement learning algorithm and the impact of the pod storage and transfer policies on the productivity of RMFS.

41 - A Bi-Level Service Capacity Planning Model with Interdependent Mdp Followers

Seokwoo Kim, Pohang University of Science and Technology, Pohang-si, Korea, Republic of, Dong Gu Choi, Yuin Moon

This talk presents a bi-level service capacity planning model with sequential-decision-making followers. This model has an important application to on-demand platform's planning problem on a service network. In the model, followers (servers) are homogeneous and their problem is formulated as constrained semi-markov decision process (SMDP) model, where reward and resource constraint budget are determined by a leader's integrated compensation and staffing problem. Further, we consider dependent followers whose interactions are reflected on sojourn time, which becomes another decision variable. Extending on a fluid-based stylized equilibrium model developed in the literature, we propose a structured bi-level program model which admits a mixed-integer bi-linear single-level reformulation. Then we derive a novel mixed-integer linear reformulation exploiting problem specific insight, along with useful valid inequalities. Further, we identify an interesting relationship between our model and network-design models. Based on this observation, a dantzig-wolfe decomposition-based solution method is developed.

42 - The Effect of Interaction Quality on Customer Satisfaction Across Response Methods: The Mediating Role of Experience Quality (EXQ)

Seonghyeon Han, Korea University, Sejong, Korea, Republic of, JAESEOK YUN, Sung Yeon Kim, Jinmin Kim

With the advancement of digital technologies and the proliferation of contactless services, organizations are increasingly adopting diverse customer interaction modes. As a result, the way customers experience and evaluate services is also evolving. This study aims to compare how different customer interaction modes—human agents, generative AI-based chatbots, and automated response systems (ARS)—affect customer satisfaction. In particular, it examines whether experience quality (EXQ), defined as the customer's emotional and holistic perception of service experience, mediates the relationship between interaction quality and satisfaction.

Interaction quality is measured through responsiveness, reliability, and communication quality, while EXQ consists of product experience, outcome focus, moments of truth, and peace of mind. A survey was conducted for each interaction type, and structural equation modeling was employed to analyze how each mode influences perceived service quality and behavioral intention. The mediating role of EXQ was also statistically tested.

This study contributes to the theoretical understanding of how technology-driven service interactions shape customer experience mechanisms. Additionally, it provides practical implications for designing interaction strategies in service environments where accuracy, consistency, and adherence to standardized procedures are critical.

43 - Constructing Profitable Master Surgery Schedules Using Column Generation-Based Primal Heuristics

Samuel Bakker, Ghent University, Ghent, Belgium, Broos Maenhout, Babak Akbarzadeh

In this research, a column generation-based math heuristic is developed that is used to solve the case mix planning problem (CMPP) whilst creating a master surgery schedule (MSS) in an integrated manner. Current literature often still treats these two as separate problems, that are solved in a sequential manner. In our study, we solve these problems in an integrated fashion. The proposed math heuristic is used to solve this problem by considering the most relevant resources in a hospital such as wards and medical staff. The CMPP focuses on finding a mix of DRGs to target that maximizes a hospital's profits, whereas the MSS allocates operating room time among surgical disciplines. The constructed MSS accounts for the amount of required operating room time for each discipline, surgical discipline preferences, and potential setup-times when two different kinds of procedures are conducted after each other in the same operating room. The developed math heuristic iteratively generates new columns, improving the incumbent. Computational experiments showcase the performance of the math heuristic against diving heuristics, the restricted master heuristic, and an MIP.

44 - Examining Opinion Dynamics on Social Media: An Agentic Simulation

Atreyi Kankanhalli, National University of Singapore, Singapore, Singapore, Prakash Sukhwal, Pratyush Yadav

The rise of social media platforms has profoundly altered public discourse and opinion formation, offering new avenues to understand social dynamics and collective behaviors. Online platforms such as X serve as forums where diverse viewpoints collide, leading to potential risks such as polarization, misinformation spread, and fragmentation. Understanding the nature of opinion evolution on such platforms is crucial for mitigating the negative consequences.

Traditional opinion dynamics models in the literature struggle to capture the complexity of real-world conversations. These models are often limited by simplifying assumptions, failing to represent the nuances of realistic conversations. Furthermore, earlier agent-based models are typically unable to account for conversational contexts.

We address these limitations by employing large language models (LLMs) as agents in synthetic social media discussion simulations, enabling more naturalistic conversations and opinion exchanges. Particularly, we investigate the critical role of moderators—individuals who guide discussions and manage information flow. Additionally, network structure properties govern information dissemination patterns across different topologies. Another salient determinant of the discourse is the nature of the discussion topic.

Our methodology manipulates these three key factors: (1) moderator presence, (2) underlying network structure (scale-free versus small-world), and (3) discussion topic nature. Through this approach, we address questions of : How does the presence of moderators influence opinion change? How does network structure affect opinion dynamics? How do different topics impact opinion change?

This study offers research contributions and insights into fostering healthier social media environments by uncovering the interplay between moderation, network structure, and topic characteristics in shaping opinion change.

45 - Recurring Recalls and the Decision to Reshore: A Behavioral Theory of the Firm Perspective

Di Liu, University of Bristol, Bristol, United Kingdom, Minhao Zhang, Di Fan, Bangdong Zhi

Reshoring antecedents' literature recognized the essential role of quality issues in driving firms' reshoring behaviors; yet, it largely failed to consider the temporal conditions and persistence of quality underperformance. Through the lens of the behavioral theory of the firm (BTOF), we argue that the persistence of quality underperformance contributes to shaping firms' reshoring decisions. We test our theory by empirically examining the relationship between the persistence of product recall events and firms' reshoring behaviors based on the panel data of 479 publicly traded U.S. firms from 2004 to 2021. We find that the probability of firms engaging in reshoring initially decreases over a period of 3.997 years and then increases as the persistence of product recall crisis persists beyond this threshold. We further find that the U-shaped relationship becomes flatter when firms perform strongly financially but steeper when they experience financial struggles. Our study contributes to reshoring and BTOF literature by exploring the role of quality underperformance persistence in inducing firms' reshoring behaviors. Our study also adds to BTOF studies by examining the interaction between financial attainment discrepancy and the persistence of quality underperformance in affecting firms' problemistic searches. Our findings indicate that managers should shed light on how the length of quality issues conditions firms' responses and their possible nonlinear relationship underlined in decision-making.

46 - Operations Research as the Backbone of Intelligent Disaster Management

Nasrin Asgari, Anglia Ruskin University, Cambridge, United Kingdom, Jihee Kim, MD Shah Newaz, Reza Zanjirani Farahani

The increasing complexity and frequency of disasters have driven the evolution of disaster management into an interdisciplinary domain, integrating OR with emerging intelligent technologies. This study builds upon a comprehensive review by Gupta et al. (2016)*, which mapped disaster management research through OR methods such as mathematical programming, simulation, and network optimization. Recent advances have seen OR frameworks synergize with Artificial Intelligence (AI), big data analytics, Geographic Information Systems (GIS), Unmanned Aerial Vehicles (UAVs), social media, augmented/virtual reality (AR/VR), deep learning, cloud computing, robotics, and digital strategies. This integration enhances decision support across all disaster phases—prevention, preparedness, response, and recovery.

In prevention and mitigation, AI-driven forecasting and GIS-based spatial analytics inform OR-based risk models and infrastructure planning. Preparedness benefits from simulation and optimization, enriched by AR/VR for immersive training. During response, real-time data from drones, sensors, and social media are incorporated into OR-driven dynamic routing, resource allocation, and decision-support systems. Recovery phases apply OR models to optimize logistics, infrastructure restoration, and resource distribution, supported by big data and cloud computing.

This synthesis positions OR not only as a set of analytical tools but as the core integrative framework that structures, optimizes, and enhances intelligent decision-making in disaster environments. By integrating both data-driven AI insights and optimization-based decision models, OR ensures that disaster management systems are adaptive, efficient, and effective in real-time, complex scenarios. The study demonstrates the importance of continuous collaboration between OR researchers, technology developers, and disaster response practitioners.

* <https://journals.sagepub.com/doi/abs/10.1111/poms.12591>

47 - Data-Driven Chance-Constrained Mixed Integer Nonlinear Bilevel Optimisation via Copulas

Syu-Ning Johnn, University College London, London, United Kingdom, Hasan Nikkhah, Meng-Lin Tsai, Styliani Avraamidou, Burcu Beykal, Vassilis Charitopoulos

Process supply chains consist of planning and scheduling as two integral components. The presence of challenges in real-world demand data, such as multivariate data dependency and correlations, noisy data distributions, and parameter uncertainties can significantly complicate the decision-making process, making it difficult to identify feasible solutions. This necessitates the adoption of data-driven optimisation approaches to effectively account for the dependency structures inherent in the observed data to efficiently explore solution spaces and identify high-quality outcomes.

This work proposes a data-driven chance-constrained data estimation framework that integrates chance-constrained programming (CCP) with copulas. CCP is an optimisation-based method enforcing stochastic constraints under specified risk thresholds. Copulas is a non-parametric technique designed to estimate data while capturing the dependence structure among multiple uncertain parameters with inherent correlations. The proposed framework leverages the strengths of both methods to achieve high accuracy in modelling variable dependencies across different scenarios, especially when data exhibit complex distributions or non-trivial dependencies.

In this work, we focus on three integrated case studies originated from the planning and scheduling of process supply chains and crude oil scheduling. We integrate our proposed estimation framework with the Data-driven Optimisation of bi-level Mixed-Integer NONlinear problems (DOMINO) framework, which is a data-driven grey-box algorithm for addressing bilevel optimisation problems, to derive decisions with guaranteed demand satisfaction rates. Computational experiments demonstrate that our proposed copula-based chance-constrained optimisation framework can incorporate demand correlation and achieve higher joint demand satisfaction rate, lower total costs with higher efficiency.

48 - Counterfactual Recourse Fairness and Feasible Actions with Causality

Shiqi Fang, University of Edinburgh, Edinburgh, United Kingdom, Zexun Chen, Jake Ansell

Machine learning is widely used in credit risk assessment for its efficiency, accuracy and generalizability, but interpretability and fairness concerns have become critical. Most counterfactual-fairness methods focus on prediction and ignore applicants denied credit—those who most need actionable recourse—while Operational Research approaches rely on purely statistical classifiers that lack causal structure. Consequently, their counterfactual explanations neither respect feature causality nor translate into real-world decision processes. We introduce a novel causal framework that unifies counterfactual fairness with interpretable recourse, asking: if a rejected female were counterfactually male, would she need less effort to gain approval, and can we offer a more diverse, feasible and equitable set of actions? Our model embeds a structural causal model, beyond minimizing recourse cost, and optimizes the number and diversity of actionable changes. To guard against biases from measurement error and model misspecification, we employ multiple generation techniques—Gaussian methods, conditional variational autoencoders and other state-of-the-art approaches—enabling robust performance in deterministic or uncertain causal settings. Supporting both distribution- and point-based counterfactuals for discrete and continuous data, our framework is especially suited to credit applications. Experiments on multiple credit scoring and synthetic datasets demonstrate consistent gains in fairness, interpretability and robustness, empowering rejected applicants with transparent, equitable pathways to approval.

49 - Learning to Trust OR Trusting to Learn? Transparency and Mutual Adaptation in Human-AI Diagnostic Collaboration

Xinyu Li, Durham Business School, Edinburgh, United Kingdom, Li Ding, Yanlu Zhao, Shuyang Wu, Timothy Kendall

This study investigates how clinicians interact with AI explanations in real time, and whether mutual learning—adaptation by both humans and AI—can emerge through repeated collaboration. Specifically, we examine how different formats of AI transparency influence diagnostic decisions, trust, and cognitive engagement in a high-stakes medical context. While many clinical AI tools prioritize accuracy, few studies explore how clinicians interpret and use AI outputs, particularly when explanations vary in complexity. We conduct a within-subject, multi-phase experiment with practicing pathologists. The study compares behavior across three transparency conditions: (1) AI prediction only, (2) prediction with key diagnostic markings, and (3) prediction with chain-of-thought reasoning. We track decision changes, interaction timing, and reasoning engagement to examine whether more explanation leads to better adoption—or cognitive overload. In a follow-up phase, clinicians evaluate the same cases without AI to assess whether earlier exposure to AI reasoning influences their independent diagnostic behavior. Findings from this study will provide empirical evidence on how to balance transparency with usability, and whether mutual learning occurs between humans and AI. The results aim to inform the design of explainable AI systems that support clinical adoption, improve human-AI collaboration, and bridge the gap between technical interpretability and real-world effectiveness.

50 - Pricing in Live Streaming E-Commerce: Integrating Customer Type Analysis

Yuan Dong, Temple University, Philadelphia, PA, United States, Guohou Shan, Subodha Kumar

Live streaming e-commerce (LSE) gradually becomes the future of shopping and creates hundreds of billions of markets. Influencers can demonstrate the product and interact with potential consumers in real-time on LSE platforms, reducing product uncertainty and raising the consumers' purchasing intention. In this research, we focus on understanding how online stores alter LSE frequency and pricing decisions to boost their revenue better. To better understand consumers' decisions, we also categorize consumers into product-oriented and influencer-oriented, aiming to understand the impact of consumer heterogeneity on store decisions. Our research makes both theoretical and practical contributions. On the one hand, we enrich the prior literature by focusing on exploring more nuanced decisions that a store can make regarding the LSE setting and adoption while considering customers' heterogeneity. On the other hand, our research findings can help online stores make better decisions and strategies as for leveraging LSE to boost their product sales and revenue.

51 - Surviving and Thriving: Firm Resilience in Competitive and Uncertain Markets

Zhide Wang, Southern Methodist University, Irving, TX, United States, Yanling Chang, Nathan Yang, Alfredo Garcia

This paper develops a dynamic equilibrium model to evaluate growth resilience—the ability of a firm to consistently expand operational capacity across both favorable and unfavorable economic environments. Recognizing that macroeconomic conditions cannot be exactly observed and must be inferred through noisy economic indicators, we formulate the problem as a partially observable stochastic game (POSG), allowing firms to update beliefs about the underlying economic states and make capacity expansion decisions accordingly. We provide theoretical identification results and develop a likelihood-based estimation method to recover the model's structural parameters. Applying this framework to study the influence of partially observable macroeconomic states on the Canadian fast-food industry from 1970 to 2004, we quantify firm-level growth resilience and find that while many firms demonstrate positive profitability across macroeconomic states, the dominant firm (McDonald's) exhibits superior dual resilience—maintaining higher growth in both downturns and booms. Simulation analyses further show that firms with weaker resilience quickly lose competitive ground amid economic volatility, whereas resilient firms successfully consolidate market share. We thus conduct a Minimum Resilience Threshold Analysis to determine critical resilience levels necessary for firms to maintain competitive parity under macroeconomic uncertainty.

52 - Influencing Travel Behavior Through Information: a Game Theory Perspective

Devin Rhoads, UConn Department of Civil & Environmental Engineering, Storrs, CT, United States

This research investigates how strategically designed information, guided by game theory, can influence individual travel behavior to enhance safety, efficiency, and equity across transportation systems. Rather than relying solely on infrastructure changes or financial incentives, this work explores how real-time and context-sensitive communication can shift behavior through perception and decision-making dynamics. Applications include detouring drivers to alleviate congestion, reducing station overcrowding in multimodal systems, and enhancing compliance with speed limits through strategically placed signage and feedback systems.

By modeling travelers as decision-makers in a shared system, game-theoretic frameworks help predict how users respond to different types of information. Concepts such as signaling, equilibrium strategy, and obedience constraints are applied to understand how messages delivered through digital displays, roadside signs, or in-app prompts can shape outcomes without coercion or financial incentives.

This project offers a novel lens on transportation planning: treating information itself similarly to infrastructure elements that shape collective mobility. The ultimate goal is to develop applicable strategies for transportation agencies to influence behavior at scale, reducing risks, optimizing flow, and promoting more equitable use of shared transportation networks.

53 - Using AI and Agent Based Modeling to Bridge Academic Findings and Organizational Practicality

Pierre Andrieux, Louisiana Tech University, Ruston, LA, United States, Justin Keeler, Jacob Whitmore

This study developed an AI based system that is designed to enhance agent-based modeling (ABM). In particular, it extracts static research findings and allows for making predictions as "what-if" scenarios. The intent was to advance simulation and replication techniques (e.g., Monte Carlo, ABM) to have greater dynamism. We calibrated this computational technique with published research that comprised of regression and difference testing of blue vs. white collar workers from Turkey. Constructs evaluated in the study included person-organization fit (POF) to predict affective/normative commitment, job satisfaction, organizational identification, and intentional to leave. The AI framework derived provides a pathway to address shortfalls in research, especially allowing for more practical usage rather than scholarly work sitting idle. Arguably, our work should cause researchers to pause and consider the idea of leveraging AI to bridge micro-level psychological mechanisms with macro-level organizational outcomes. In our calibration testing, the technique shows promise to advance ABM methodologically in organizational behavior research.

54 - Policy Gradient Optimization for Markov Decision Processes with Epistemic Uncertainty and General Loss Functions

Xiaoshuang Wang, Georgia Tech, Atlanta, GA, United States, Yifan Lin, Enlu Zhou

Motivated by many application problems, we consider Markov decision processes (MDPs) with a general loss function and unknown parameters. To mitigate the epistemic uncertainty associated with unknown parameters, we take a Bayesian approach to estimate the parameters from data and impose a coherent risk functional (with respect to the Bayesian posterior distribution) on the loss. Since this formulation usually does not satisfy the interchangeability principle, it does not admit Bellman equations and cannot be solved by approaches based on dynamic programming. Therefore, We propose a policy gradient optimization method, leveraging the dual representation of coherent risk measures and extending the envelope theorem to continuous cases. We then show the stationary analysis of the algorithm with a convergence rate of

$\mathcal{O}(T^{-(1/2)} + r^{-(1/2)})$, where T is the number of policy gradient iterations and r is the sample size of the gradient estimator. We further extend our algorithm to an episodic setting, and establish the global convergence of the extended algorithm and provide bounds on the number of iterations needed to achieve an error bound $\mathcal{O}(\epsilon)$ in each episode.

55 - Reframing Airport Choice in Platform Logistics: the Case of Amazon Air

Xiaofu Fan, Purdue University, West Lafayette, IN, United States, Karanki Fecri, Chien-Tsung Lu

While the substantial growth of e-commerce has elevated the role of air cargo from a peripheral service to a critical revenue source within the air transportation industry, airports, as essential nodes in this network, have increasingly benefited from this shift. In particular, the emergence of new players such as Amazon Air has significantly influenced the revenue growth trajectories of many airports. Despite this trend, the determinants guiding Amazon's airport selection strategy remain poorly understood. By analyzing data from 62 U.S. airports between 2010 and 2023, this study employs a Instrumental Variables Probit model to shed light on the key factors that influence Amazon's airport choices, thereby providing insights to help airports better align with the operational needs of major e-commerce carriers. The results reveal that Amazon prefers to choose airports that are non-hub, have good freight infrastructure, are far from urban centers, have low ground delays, and have a large scale of regional employment, while the per capita income has no significant level. These findings indicate the core logic of the "decentralized" layout of e-commerce logistics: by avoiding the high cost and congestion of traditional hubs, relying on professional facilities and flexible scheduling to achieve efficient performance, which confirms its high sensitivity to operational efficiency and scheduling freedom.

56 - Hierarchical Planning with LLM Agents for Simulating EV Charging Behaviors Under Demand Response

Yi Ju, University of California, Berkeley, Berkeley, CA, United States, Lunlong Li, Jingchun Wang, Jinhua Zhao, Scott Moura

With the rapid global adoption of electric vehicles (EVs), unmanaged charging poses serious risks to the power grid by exacerbating demand imbalances with renewables and overloading distribution infrastructure. Conversely, coordinated EV charging could transform millions of vehicles into a distributed, virtual energy storage system, rivaling the total capacity of today's utility-scale batteries. Realizing this vision, however, hinges on effective mechanisms that align individual driver incentives with broader system benefits, without compromising personal mobility needs.

This study explores the use of large language model (LLM) agents to simulate individual EV charging behavior in response to demand response (DR) signals. We introduce a novel decision-making task grounded in realistic, microscopic mobility trajectories and geolocated charging infrastructure. The LLM agent is given a detailed

two-day activity schedule together with tools to query charging station availability and pricing. It must then generate executable plans on when, where, and how to charge accounting for feasibility constraints, diverse pricing schemes, and its assigned persona. The task presents a combinatorially large solution space, requiring the agent to perform careful calculation and forward-looking reasoning to balance multiple objectives such as cost, time, uncertainty, behavioral inertia, and social norms. Leveraging a hierarchical planning workflow based on the ReAct and Reflexion paradigms, the LLM agent demonstrates plausible planning capability. This first-of-its-kind experiment offers a new lens for modeling complex human behaviors in EV charging, which opens the door to more realistic large-scale simulations for energy-mobility systems to assess the effectiveness of demand flexibility programs.

57 - Advancing Multiple Regression-Based Resilience Predictions with Transfer Function Methods

Fatemeh Salbough, University of Massachusetts Dartmouth, Dartmouth, MA, United States, Priscila Silva, Lance Fiondella

This research introduces a more advanced time series modeling technique that incorporates transfer functions to enhance resilience model prediction. Unlike traditional statistical methods that struggle with flexibility and long-term dependency forecasting, this approach effectively tracks and predicts system performance changes when facing multiple shocks and stresses of varying intensity and duration.

The transfer function methodology provides a structured framework for planning resilience assessment tests tailored to specific disruptions while guiding efficient data collection. Though resilience engineering is domain-specific, this versatile approach can be applied across various fields.

To validate its effectiveness, the researchers compared the transfer function model with regression models using interaction terms, analyzing historical data on job losses during the 1980 U.S. recessions. The numerical results, evaluated through goodness of fit measures, demonstrate the transfer function model's superior predictive capabilities, offering a compelling alternative approach to resilience assessment.

58 - Ultra-High Dimensional Threshold Selection for Quantile Feature Screening with Fdr Error Rate Control: A Case Study on High Blood Pressure Analysis

Saidat Sanni, University of Texas at San Antonio, San Antonio, TX, United States, Yan Yu, Zhigen Zhao

Identifying important genetic risk factors for high blood pressure is paramount for targeted interventions and comprehensive management strategies for this prevalent health condition. With ultra-high dimensional genetic factors, feature screening plays a vital role in retaining the important risk factors and screening out the unimportant ones. However, its efficacy depends on selecting the appropriate threshold in different data contexts. We propose the novel Quantile Mirror (QM) approach for data-adaptive threshold selection with false discovery rate (FDR) control in quantile feature screening, where abnormally high blood pressure levels are of the most interest. The symmetric QM statistics enable the estimation of the FDR. Multiple data splitting is employed for result stability and different thresholds are allowed for different quantiles. While many of the retained high-blood-pressure-related genetic features are validated from medical literature, some new risk factors are preserved. Additionally, we introduce the Quantile Reflection via Data Splitting (QREDS) and the Hard Threshold with QREDS feature screening procedures. We show that our proposed procedures asymptotically control the FDR under suitable conditions. We further demonstrate the performance of the proposed methods through extensive numerical analysis and simulation studies.

59 - The Performance Effects of Franchisee Autonomy in Operational Decisions: Evidence from a Restaurant Franchise

Jaewoo LEE, University of South Carolina, Columbia, SC, United States, Keith Skowronski, Sean Handley

Franchise systems balance the benefits of standardization with the need for local adaptation. While most prior research has focused on customer-facing autonomy, we examine how franchisees' autonomy in three operational dimensions—supplier selection, internal process compliance, and product assortment—affects outlet-level quality and sales performance. Using detailed monthly panel data from 90 outlets of a Korean restaurant franchise (2021–2023), we find that sourcing from alternative suppliers negatively impacts quality performance. Deviations from standard recipes enhance quality but reduce sales. Similarly, narrowing the adoption of brand-standard items dampens sales without yielding quality benefits. Post hoc analyses reveal that franchisor audits can mitigate the quality risks of using alternative suppliers and that the direction of recipe deviations matters—overserving improves satisfaction but undermines sales, while underserving reduces sales without enhancing satisfaction. We also show that adopting official side dishes, rather than entrées, is more effective in driving sales growth. These findings advance the franchising literature by shifting the focus from granted or perceived autonomy to exercised autonomy and by uncovering its differential performance effects across operational domains. For managers, the results suggest the need for differentiated autonomy strategies that involve working with franchisees to align their autonomous decisions with what has proven effective for sales and quality at the outlet level. This study offers actionable insights for franchise system design, governance, and performance optimization.

60 - An Integrated Assignment, Scheduling and Routing Optimization Framework for Humanitarian Food Distribution: A Case Study of the Banco De Alimentos De Bogotá

Ahmad Torabzadeh, university at Buffalo, buffalo, NY, United States, Sabrina Casucci, Robert Dell, Gonzalo Mejia

Food insecurity continues to be a pressing global challenge, impacting millions of people worldwide. The Banco de Alimentos de Bogotá, a food bank in Colombia, serves as a crucial support system for approximately 500,000 individuals through a network of 500 partner agencies. This research tackles the complex operational challenge of distributing food donations efficiently while managing limited resources and ensuring equitable delivery workloads. We develop an integrated agency assignment and vehicle routing model tailored to the food bank's daily logistics operations over a one-month planning horizon.

We solve our model in two steps with little loss of optimality. First, an integer linear programming (ILP) model assigns agencies to specific delivery days, accounting for varying delivery frequencies and aiming to balance workload across days. Second, for each day, we solve a capacitated vehicle routing problem (VRP) with time windows, route duration limits, and homogeneous vehicle constraints that minimizes total routing costs while promoting fairness in workload distribution across the fleet.

While each food bank operates under unique logistical, geographic, and organizational constraints, our model offers a flexible and adaptable optimization framework that can be customized to fit different operational contexts. By iteratively integrating agency assignment and daily routing, the model captures real-world delivery constraints and improves efficiency in both scheduling and routing. This contributes to humanitarian logistics by providing a practical tool that can help food banks enhance their operational performance and better serve communities in need.

61 - Rethinking Multitasking: IT, Teams, and Temporal Simultaneity

Li Zhang, MICHIGAN STATE UNIVERSITY, East Lansing, MI, United States, Julie Ryan Wolf, Brian Pentland

Modern organizational work is characterized by multitasking through concurrent task execution and frequent task switching. Extant research has primarily conceptualized multitasking as an individual-level, time-varying behavior defined by temporal concurrency, so it tends to overlook the role of teams and the enabling power of information technology (IT) in multitasking. True simultaneity of task execution has been dismissed as a misnomer due to assumed human limitations. Our work shifts the focus from individual to team-level multitasking and examines how IT and team enable multitasking behavior in practice. We conceptualize multitasking as composed of three distinct forms: pure IT-enabled multitasking, pure socially enabled multitasking, and hybrid socio-technical multitasking. Using electronic health record (EHR) audit logs from the University of Rochester Medical Center (URMC), we examine the impact of different multitasking forms on task documentation burden (measured as team-EHR interaction time). Our study shows that a 10% increase in pure IT-enabled multitasking is associated with a 41% decrease in task documentation time and a 10% increase in hybrid socio-technical multitasking is associated with a 35% decrease in task documentation time. Endogeneity in the model is addressed using Heteroskedasticity-based IV-method. Our study contributes to the multitasking literature by conceptualizing multitasking as a team- and IT-enabled phenomenon grounded in temporal simultaneity. Our findings demonstrate that the multitasking affordances of IT—when effectively enacted—can improve worker well-being through reduced documentation burden. The findings offer practical implications for IT value, system design, and IT-training interventions aimed at enhancing multitasking efficacy in complex, high-tempo environments.

62 - Robust Hedging Strategies for Community Choice Aggregators: Managing Market Price Uncertainty

Manar Jaradat, State University of New York, Binghamton, Binghamton, NY, United States, Neha Patankar, James Himelick

Community Choice Aggregation (CCA) is an energy procurement model that empowers local governments to purchase electricity on behalf of retail customers, giving communities greater control over energy sourcing and the ability to prioritize cleaner, renewable options. The increasing integration of renewable energy and storage fundamentally reshapes risk management in electricity systems. While this combination enhances system flexibility and resilience, and reduces operational and financial risks, it also introduces new complexities in planning and operation. Traditional risk strategies often rely on fixed probability distributions and assumptions that may fail under high uncertainty in evolving energy markets. In contrast, robust optimization explicitly incorporates ambiguity in model parameters and worst-case scenarios, offering stronger protection against volatile market conditions. This research develops a robust hedging strategy to address market price uncertainty, specifically tailored to CCAs. By leveraging robust optimization techniques and adapting conventional capacity expansion models, we support the design of diversified energy portfolios that are both cost-effective and resilient to uncertain market conditions, aligning with the sustainability and affordability goals of CCAs. The study provides two key contributions: First, insights into how renewable energy and storage integration transform the effectiveness of traditional risk management strategies. Second, practical recommendations for CCAs to enhance resilience and cost efficiency through robust portfolio planning. These findings support more informed decision-making for local energy programs and contribute to the broader effort of transitioning to a reliable, low-carbon energy future.

63 - Course Scheduling at Florida Polytechnic Using Genetic Algorithms

Cindy Ceijas, Florida Polytechnic University, Lakeland, FL, United States, Alexander Joyce

School operations, such as the timetabling problem, become a difficult optimization model as school enrollment increases. Allocating resources such as educators, classrooms, and time slots, under criteria like course level and course sequencing, is an integer NP-hard problem. Genetic algorithms provide an alternative to traditional optimization techniques by simulating evolution and natural selection over the set of possible solutions. This poster examines the use of genetic algorithms to solve the timetabling problem for Florida Polytechnic University. Our analysis investigates different heuristics involving crossovers using multiple fitness indicators, and variable mutation strategies using different genetic algorithm parameters.

64 - Optimization Models for the Ada Paratransit Districting Problem

Yanshuo Sun, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Chinonso Okorie, Shijie Chen

Small and medium-sized paratransit operators typically provide services within a centralized region. By contrast, larger operators often divide their service areas into multiple zones, each managed by an independent contractor. Although this zoning strategy offers benefits, such as improved vehicle on-time performance, the zonal boundaries pose significant barriers to inter-zonal passengers. Despite the wide adoption of the multi-zone structure in major U.S. metropolitan areas, there is a notable research gap in the literature: no rigorous quantitative methods exist to support those important paratransit districting decisions. To address this gap, we present two integer programming models to assist transit authorities in dividing a paratransit service region into two zones. The optimization objective is to minimize inter-zonal trip requests caused by the established boundary while ensuring some workload balance. As the first model serves motivational purposes and is a simplified version of the second, we thoroughly tested the latter model using both synthetic and real-world data. The promising results suggest that the designed model could significantly advance paratransit districting practices for major U.S. paratransit authorities.

65 - A Comparative Evaluation of Transformer and Mamba Models for Healthcare Time-Series Applications

Abigail Binder, Rochester Institute of Technology, Rochester, NY, United States, Nasibeh Azadeh-Fard

Transformers have become the de facto standard for sequence modeling, while Mamba—a newly proposed state-space model—offers a compelling alternative that remains largely underexplored. In this poster, we conduct a structured literature-based comparison of both architectures, evaluating them along three key dimensions: computational complexity, interpretability, and data efficiency. Drawing on studies in healthcare time-series (e.g., patient vital signs, length-of-stay forecasting), we identify which model characteristics align best with resource-constrained clinical settings.

66 - An MDP Model of Inventory Management Policies for Managing Drug Shortages in Hospital Pharmacies Under Uncertain Lead TIME Information

Noah Chicoine, Northeastern University, Boston, MA, United States, Jacqueline Griffin

According to the ASHP, there are hundreds of drug products currently on shortage across hospitals in the United States, and that number is increasing. With the prevalence of drug shortages, shortage management has become an ongoing practice at hospital pharmacies. Hospitals employ a number of strategies to mitigate shortages, including decreasing utilization in certain floors, encouraging use of substitutable products, and ordering alternative products. Additionally, pharmacies utilize “estimated release dates (ERDs)”, a form of lead time information from upstream suppliers, to determine which mitigation strategies may be best to implement with respect to costs and impact on patients. Research has shown that ERDs can be inaccurate and subject to change before shipments arrive at the hospital. To investigate how this uncertain lead time information effects optimal inventory management decisions, we developed an MDP model of pharmacy decision-making during a drug shortage. By including actions that mimic extending product utilization and switching to alternative medications, we test the effects of different system costs on optimal policies, showing how different types of products may require vastly different management strategies.

67 - Filling Gaps OR Replacing Creators? Strategic AIGC Adoption in Stock Content Platforms

Xiaoxiang Wang, Tsinghua University, Beijing, China, People's Republic of

AI-generated content (AIGC) allows consumers to create customized content, challenging stock content platforms whose revenues depend on licensing professional user-generated content (PUGC) to consumers. However, under the scaling law of AI training, large PUGC inventories are valuable assets for launching platform-owned AIGC. Introducing AIGC can fill gaps in existing contents and generates new revenue, but may also substitute for PUGC, risking creator exit. We develop a two-period game-theoretic model to explore whether a stock content platform should introduce AIGC trained on creator contents, and how should it adjust revenue shares with creators. Unlike prior work, our model uniquely captures the dependency of AIGC quality on PUGC training data volume and AIGC’s generalization capability. Our results show that AIGC adoption is always optimal for small markets, but for large markets, profitability hinges on AI learning rates exceeding a critical threshold. Contrary to common belief, adoption does not necessarily lead to lower revenue shares for creators—platforms may raise shares either to address initial UGC shortages (small markets) or to accelerate AI training (high learning rates). Welfare analysis reveals that AIGC adoption can either benefit or harm all stakeholders depending on market parameters. Our results provide strategic guidance for platform operators on balancing new AI technology adoption with creator ecosystem sustainability.

68 - Hierarchical Data Modeling: a Systematic Comparison of Statistical, Tree-Based, and Neural Network Approaches

Marzieh Amiri Shahbazi, Rochester Institute of Technology, Rochester, NY, 14623, NY, United States, Nasibeh Azadeh-Fard

Hierarchical data structures are common in healthcare and other domains, yet rigorous comparisons across distinct modeling paradigms remain limited. This study presents a systematic evaluation of three hierarchical approaches—statistical (Hierarchical Mixed Models), tree-based (Hierarchical Random Forest), and neural (Hierarchical Neural Networks)—to model nested healthcare data. Using the 2019 National Inpatient Sample comprising over 7 million records from 4,568 hospitals across four U.S. regions, models were assessed in predicting length of stay (LOS) across patient, hospital, and regional levels. A comprehensive evaluation framework integrated both performance metrics and structural analysis. Model robustness was tested across different sample sizes, simplified hierarchies, and a separate intensive care dataset (MIMIC-IV). The tree-based approach consistently outperformed others in predictive accuracy, variance explanation, and stability, while maintaining computational efficiency. Neural networks captured mid-level structure more effectively but required greater computational resources and exhibited systematic bias. Statistical models offered fast inference and interpretability but struggled with prediction accuracy at intermediate levels. These findings provide practical guidelines for selecting appropriate hierarchical modeling methods based on data scale, computational constraints, and the nature of nested structures. This work contributes a unified benchmark and highlights trade-offs in scalability, interpretability, and precision—guiding researchers and practitioners toward informed methodological choices in hierarchical learning tasks.

69 - Blockchain Implementation in Internet of Things (Iot) Based Sustainable Supply Chain: Development a Framework Using Fuzzy Entropy-Waspas

Atefeh Shoomal, University of Texas at Arlington, Arlington, TX, United States, Mohammad Jahanbakht, Paul Componation, Paul Componation

The deployment of Internet of Things (IoT) and blockchain technologies plays a critical role in enhancing efficiency, resilience, and sustainability in supply chains. Despite their potential, adoption challenges complicate strategic decision-making. Selecting the optimal blockchain-IoT integration use case mitigates business risks, yet this decision is complex and influenced by multiple criteria. This study identifies key factors affecting blockchain-IoT adoption in sustainable supply chains through a comprehensive literature review based on the Technology-Organization-Environment framework. We propose an Entropy-Weighted Aggregated Sum Product Assessment (WASPAS) framework under a q-rung orthopair fuzzy set to prioritize these criteria and evaluate blockchain-IoT use cases for implementation in sustainable supply chain. The model integrates expert preferences and is validated through a manufacturing supply chain case study. The result of case study shows technology maturity, sustainability, and investment cost as top ranked criteria and IoT tracking with blockchain as the optimal use case. Sensitivity and comparative analyses further demonstrate its robustness and effectiveness. By reducing uncertainty, this model empowers supply chain decision-makers to adopt blockchain-IoT solutions that enhance sustainability and operational efficiency.

Keywords: Blockchain; Internet of Things; Sustainable Supply Chain; Technology-Organization-Environment; Entropy-WASPAS; q-ROFSs

70 - Prediction of Road Accident Severity Using Multimodal Machine Learning and Large Language Models

Felipe Villa, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Traffic accident-severity prediction has historically treated numeric traffic signals and crash narratives as disjoint data silos. We introduce a multimodal pipeline that unifies minute-level flow measurements from Automatic Vehicle Identification sensors with contextual embeddings extracted by a large language model from 2.070 historical crash reports. Central to our formulation is Medical-Assistance-Required Accident (MARA), an ordinal target that bridges engineering and clinical priorities by explicitly encoding whether on-scene medical support will be needed.

Sensor features and text embeddings are fused through cross-modal attention and fed to a gradient-boosted classifier. The resulting model achieves 94.4 % accuracy, 94.6 % precision, and 94.5 % F1, exceeding the best unimodal baselines by ≥ 20 percentage points. Most notably, recall for non-severe accidents leaps from 58 % to 81 %, while MARA recall reaches 93 %, ensuring that events demanding urgent care are rarely overlooked.

To our knowledge, this is the first study to deliver real-time accident-severity forecasting at motorway scale via sensor-language fusion. By coupling the new MARA variable with large-language-model embeddings, the framework not only advances predictive performance but also supplies a directly actionable triage layer for traffic-control centres. These findings establish a new benchmark for multimodal safety analytics and demonstrate the transformative potential of integrating foundation-model semantics into intelligent transportation systems.

71 - From Formulation to Programs are LLMs Ready for Operation Research?

Guoyun Zhang, National University of Singapore, Singapore, Singapore

The integration of Large Language Models (LLMs) with optimization modeling offers a promising avenue for advancing decision-making in operations research (OR). Traditional optimization methods—such as linear programming, mixed integer programming, and simulation—depend heavily on domain expertise to translate real-world problems into solvable mathematical models. While solvers like Gurobi and COPT are powerful, expert input remains essential for defining objectives, constraints, and variables.

This research investigates the potential of LLMs, specifically the DeepSeek-R1 model, to bridge this formulation gap using natural language understanding and code generation. Although prior models like GPT-4, Claude, and Bard have shown strong performance in NLP and reasoning tasks, their high token costs and tendency toward hallucinations limit real-world applicability in supply chain contexts. In contrast, DeepSeek-R1, a cost-efficient and high-performing model trained with reinforcement learning, presents a viable alternative. Despite its success in benchmarks such as LiveCodeBench and Math-500, its effectiveness in applied OR scenarios remains under explored.

This study systematically evaluates DeepSeek-R1 across four key OR benchmarks: NL4OPT, IndustryOR, EasyLP, and ComplexOR. Our methodology includes baseline assessments, the development of a hallucination taxonomy, and the application of mitigation strategies like LLM-as-a-Judge, Few-shot Learning (FSL), Tool Calling, and a Multi-agent Framework. These techniques aim to reduce hallucinations, enhance formulation accuracy, and better align model outputs with user intent.

72 - Optimal Targeted Intercity Travel Restriction Policies

Liam Ouyang, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Xu OUYANG, Sai-ho Chung, Tsan-Ming Choi

As an effective non-pharmaceutical tool, travel restrictions are often adopted to contain the early spread of severe infectious diseases (e.g., SARS, COVID-19, and monkeypox). This study examines targeted intercity travel restriction policies to hedge against potential pandemic deterioration. We consider the correlation of disease spread among cities in deteriorating scenarios, modeled by distributionally robust optimization (DRO). The problem is formulated by a nonlinear huge-scale DRO model and then reduced to a mixed-integer linear equivalence of moderate scales. We develop a Benders decomposition algorithm, utilizing a special relation between master- and sub-problem variables. Experiments show our model “mean-variance dominates” a classic robust optimization model that ignores correlation of spread, in terms of higher pandemic control performance and lower performance variation. In the case study, targeted policies exhibit two to five times the pandemic control performance (i.e., infection reduced or human lives saved) against uniform travel restriction policies adopted in the real world. Policymakers are advised to impose targeted intercity travel restrictions as early as possible, since pandemic control performance is more obvious at the early outbreak stage. Targeted restriction policies are promising for developing countries because of comparable pandemic control performance under limited budgets, compared with those with adequate budgets.

Monday, October 27, 2:45 PM - 4:00 PM

MD01

Building A Level 3 A301

Nextmv/Hexaly

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - DecisionOPS in the Analytics Process: Collaboration, Observability, Testing, Ecosystem, and Workflows

Carolyn Mooney, Nextmv, Philadelphia, PA, United States, Tiffany Bogich

Optimization is founded upon the promise of efficiency and improving solutions. In today’s analytics environment, the leading data science and optimization teams will be defined by how they apply DecisionOps practices to test, deploy, and operate these solutions alongside software and business stakeholders. Join this session to learn basic concepts, real-world applications, and how to get started with DecisionOps in Nextmv, a platform for accelerating decision model development. This session is relevant to anyone working with tools such as: OR-Tools, Pyomo, HiGHS, VROOM, Gurobi, AMPL, FICO Xpress, Hexaly, Jupyter Notebooks, Statsmodels, Prophet, scikit-learn, Databricks, Snowflake, and more.

2 - Computing Dual Bounds of Set-Based Models Using Column Generation and Column Elimination in Hexaly

Julien Darlay, Hexaly, Brooklyn, NY, United States

Hexaly is a model-and-run solver that integrates heuristics and exact methods. A set-based modeling formalism was introduced to simplify the modeling of some combinatorial problems, like routing or packing problems. For instance, in a routing problem, list variables can be used to model the sequence of visits made by each truck. These decision variables are well suited for a heuristic search but are much more challenging to integrate into a mathematical programming approach to compute

lower bounds. A direct MILP reformulation introduces a quadratic number of binary decisions with several big M constraints, leading to poor scalability and bounds. Hexaly automatically detects such structures in a user model and reformulates them in an extended MILP model to compute lower bounds parallel to a heuristic search. This model is solved efficiently using the literature's state-of-the-art branch-and-cut-and-price techniques and column elimination algorithms. This talk will present the general approach, the algorithms used for the resolution, and some benchmarks on the classical vehicle routing and packing problems.

MD02

Building A Level 3 A302

Panel on Quantum Computing and OR

Panel Session

Committee's Choice

Co-Chair: Tamás Terlaky, Lehigh University, 200 W. Packer Avenue, Bethlehem, PA, PA, 18015, United States

1 -

Tamás Terlaky, Lehigh University, Bethlehem, PA, PA, United States

2 - Panelist

Carleton Coffrin, Los Alamos National Laboratory, Los Alamos, NM, United States

3 - Panelist

Stefan Woerner, IBM, Zurich, Switzerland

4 - Panelist

Sridhar Tayur, Carnegie Mellon University, Pittsburgh, PA, United States

5 - Panelist

Ojas Parekh, Sandia National Labs, Albuquerque, NM, United States

6 - Panelist

David Bernal, Purdue University, Lafayette, IN, United States

MD03

Building A Level 3 A303

Methodology and Applications in Group Decision and Negotiation

Invited Session

Group Decision and Negotiation

Chair: Danielle Costa Morais, Universidade Federal de Pernambuco, Recife, N/A, Brazil

1 - Behavioral Asymmetry in the Newsvendor Problem: a Fuzzy Prospect Theory Approach

WenYi Lee, National Taipei University of Business, Taipei, Taiwan, Yu-Rong Chen

The newsvendor problem is a foundational model in inventory and revenue management, traditionally assuming rational decision-making and risk-neutral preferences. However, real-world decisions often reflect asymmetric psychological responses to overordering and underordering. This study addresses this behavioral complexity by applying fuzzy prospect theory to the newsvendor framework. Our model integrates fuzzy logic to capture ambiguity in human perception and prospect theory to reflect loss aversion and reference dependence in inventory decisions. By accounting for the cognitive differences in how decision-makers perceive surplus versus shortage, our approach offers a more realistic understanding of inventory behavior under uncertainty. Numerical experiments demonstrate that incorporating behavioral asymmetry can significantly alter inventory strategies, with implications for more nuanced revenue management practices.

2 - Carbon Emission Reduction in Manufacture-Circulation Industrial System: An Assortment Optimization Embedded Cooperative Game Theory Method

Xihua Liu, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, China, Shenyang, China, People's Republic of, Shengnan Zhao, Lixin Tang

In recent years, significant carbon emission from human activity has led to global climate change, and many countries have begun to take measures to address this severe situation. In 2020, China proposed the goal of carbon peak and neutrality, which aims to peak carbon emission before 2030 and achieve the balance between carbon emission and absorption by 2060. In response to this, it is necessary for vertically aligned firms to collaborate to integrate resources and reduce repetitive production. This study focuses on the cooperation between firms producing steel, components and vehicles. A cooperative game theory model is adopted. Firms within the same product category can form coalitions to share their technology and jointly undertake the cost of production and carbon emission. The value of characteristic functions is determined by an assortment optimization model, where firm coalitions can choose a subset of the products they can produce to exhibit downstream firms. The objective is to minimize the total cost of the firm coalition. We employ a row generation framework: Solve the assortment optimization problem to obtain the characteristic function values of certain coalitions, and use row generation to tackle with the computational difficulty caused by the exponential number of possible coalitions. Different methods are used to allocate the total cost of a coalition to its members in a comparative way. Numerical experiments show that the formation of coalitions between firms reduces total carbon emission when compared with the non-cooperative scenario. The results provide insights for low-carbon collaborative production management between firms.

MD04

Building A Level 3 A304

AI, Sentiment, and Behavioral Drivers in Financial Markets

Contributed Session

Contributed

Chair: Joomi Park, Yonsei University, Seoul, Korea, Republic of

1 - Evolutionary Model of Data Market

Kewei Cai, Beihang University, Beijing, China, People's Republic of, Junhuan Zhang

The rapid expansion of data markets has garnered increasing scholarly and industry interest. We propose an evolutionary model of a data market microstructure including consumers, producers, and data platforms. Data platforms act as intermediaries, offering consumer data services and digital advertising services to facilitate product transactions. They collect consumption data from consumers and reveal it to help producers make informed decisions. Likewise, producers can advertise their products through these platforms. Consumers decide whether to share their data, while producers choose whether to advertise. Data platforms determine if they will provide data services. By examining the evolutionary strategies of each participant, we identify the equilibrium where all participants actively engage in the data market. This equilibrium is stable under certain pricing conditions and is influenced by several key factors: 1) Data platforms must balance pricing conditions between consumer data and digital advertising services, ensuring prices cover the opportunity costs of all participants. 2) Consumers' willingness to share data depends on privacy loss, search

cost, and compensation fees. 3) Producers' willingness to advertise is influenced by production cost, service costs, advertising impact, and product price. 4) Data platforms' willingness to provide services depends on data production cost, collection cost, and service costs. Also, we extend the model by incorporating dynamic incentives, pricing structures, and personalized advertising to enhance data sharing and advertising willingness. This study provides insights into data asset pricing and data market design.

2 - The Effect of Selling Asset Behavior on Investors' Returns: Analysis Using Data from Robo-Advisor Users

Norio Hibiki, Keio University, Yokohama, Japan, Yusuke Ushijima

Long-term investment and diversification are crucial for individual investors who employ asset management strategies using mutual funds and robo-advisors. However, some investors sell or withdraw funds for various reasons. These trading behaviors affect actual investor returns. Morningstar (2024) has demonstrated the investor gap between the total return of funds and the investor return, which is caused by trading behavior. Barber and Odean (2000) analyzed individual trading data and showed that frequent trading leads to lower performance, which has an unfavorable effect on asset growth. Therefore, this study focuses on investors' withdrawal behavior and analyzes the difference in investor returns and the relationship between them. We analyze transaction records (data) from approximately 380 thousand users of WealthNavi, a robo-advisor service provided in Japan. While results vary depending on the evaluation period, we find that users who do not sell their portfolio and invest over a longer period tend to have higher investment returns, on average, than those who have sold their portfolio. A regression analysis was conducted with the investor return as the dependent variable, a dummy variable representing selling asset behavior as the independent variable, and dummy variables for investment experience, additional deposits, investment dates, and valuation dates as control variables related to investor returns. The results showed that the selling asset behavior has a statistically significant negative correlation with investors' returns.

3 - Post COVID-19 Impact on Remote Work on the Housing and Commercial Real Estate

Victor Dichi, University of Miami, Miami, FL, United States, VINCENT OMACHONU

Since the onset of COVID-19, the amount of employed people working from home in the United States rose from 5.7% in 2019 to a staggering 17.9% in 2021 (Espitia et. al 2021, US Census Bureau 2022, Laura Pop-Badiu et. al 2024). As a result, the use of office space has declined dramatically, and the amount of office space still on the market, as well as the offices currently being built are not necessarily optimal. Due to this surplus of offices, the demand for office space has decreased, causing there to be a decrease in the efficiency of the office market. People now need more housing and housing space for their work, as people now need more space in their homes to work. Consequently, vacancy rates in the housing market have decreased, causing an incredible increase in housing demand, therefore creating a shortage in the market (Guirguis et. al 2023). The purpose of this research is to analyze the economic impact of remote work on the use of office buildings as well as housing real estate in major cities in the United States. Additionally, the study seeks to propose a way to stabilize this crisis through the use of economics and mathematical approaches.

4 - Incorporating News Headline Sentiment into the Markowitz Portfolio Selection Model

Joomi Park, Yonsei University, Seoul, Korea, Republic of, Seongmoon Kim

This paper proposes a method to enhance Markowitz's mean-variance portfolio model by incorporating sentiment indexes derived from news headline data. Specifically, the model adjusts the portfolio weights of individual stocks based on the results of sentiment analysis. The sentiment data were collected from the headlines of ten major Korean economic newspapers. To evaluate the effectiveness of the proposed model, we conducted a backtest using historical monthly stock return data from the South Korean stock market over a ten-year period, from March 2015 to February 2025.

MD05

Building A Level 3 A305

INFORMS Ambassadors

Invited Session

Ambassadors Program

Chair: Banafsheh Behzad, California State University, Long Beach, Long Beach, CA, United States

1 - Well Informed: Increasing Wellness Capacity and Community Within Academia

Jessye Talley, Morgan State University, Baltimore, MD, United States

The Well INFORMED Mentorship Program provides space for faculty and PhD students to take a holistic approach to their academic experience. By understanding the intersection between their values, identity, and wellness they able to make connections that can inform their academic life. This talk will highlight the nuances and outcomes of this program.

2 - Levering the INFORMS Annual Meeting Host City Rotation to Improve the Visibility of Orms Among Students at Puis: the Inform's Promise Scholars Program

Juan Martinez Mori, University of Colorado Denver, Denver, CO, United States

One of the challenges with making ORMS more accessible starts with its name: compared to other STEM fields, the terms "Operations Research" and "Management Science" are not particularly indicative of what they are all about. In this INFORMS Ambassadors project, we leverage the Annual Meeting host city rotation to perform targeted outreach at local primarily undergraduate institutions (PUIs) to involve students in the Annual Meeting and introduce them to career opportunities in ORMS.

3 - A Machine Learning Approach for Identifying Marginalized Subgroups of INFORMS Members and Participants

Matt Baucum, Oklahoma State University, Stillwater, OK, United States

In this project, we analyze previously-collected survey data of INFORMS members and non-members to better understand the professional and demographic characteristics of INFORMS, how such characteristics have changed over time, and how such characteristics contrast with similar professional and academic organizations. Importantly, we also conduct a machine learning-driven heterogeneity analysis to identify subgroups of individuals who report feeling marginalized or excluded through participation in INFORMS events, and investigate whether such experiences serve as a barrier to membership for academics or professionals.

Data come from two INFORMS surveys administered in 2022 (to INFORMS members) and 2024 (to members and non-members), with several thousand completed survey responses across both years. We use factor analysis to identify dimensional scales that reliably quantify members' overall satisfaction with INFORMS and satisfaction with INFORMS' inclusivity. We then employ interpretable machine learning techniques to identify subgroups of individuals who are particularly likely to report negative experiences with INFORMS, or particularly unlikely to place membership. We also combine such techniques with autoregressive analysis to assess whether particular groups experienced notable changes in INFORMS satisfaction or membership intention between the 2022 and 2024 surveys.

4 - Operations Research in Women's Health and Healthcare: a Scoping-Inspired Narrative Review and Roadmap for the Future

Rachel Stephenson, University of Toronto, Toronto, ON, Canada, Rachel Wong

Operations Research (OR) approaches have been applied to derive meaningful policy insights and effective solutions to a broad range of healthcare problems, but we believe that there is a gap in OR literature addressing issues that differently or disproportionately impact women's health and healthcare (hereafter referred to as "issues in women's health and healthcare" for brevity). As such, we are undertaking a project to identify this gap and promote a research agenda to close it. The project consists of three phases:

Phase 1: Perform a scoping-inspired review of literature on significant issues in women's health and healthcare studied and published in top OR journals over the past ~10 years.

Phase 2: Distribute a survey to experts in women's health and healthcare with the research gaps identified in Phase 1. Experts will be asked to prioritize research areas, identify specific subproblems within broad areas that may be solved using OR approaches, and identify any emerging issues that were excluded from the list provided.

Phase 3: The results from Phases 1 and 2 will be synthesized and disseminated in a survey to the OR community. Experts will be asked to describe how OR could be applied to solve these problems.

The primary outcome of this work will be the identification of opportunities for novel applications of OR in women's health and healthcare. The goal of this project is to promote applications of OR to women's health and healthcare issues, and therefore, to ultimately improve women's health outcomes.

5 - Driving Engagement with Data: a Predictive Analytics Challenge for Community Impact

Imtiaz Ahmed, West Virginia University, Morgantown, WV, United States, Nathan Gaw

This project launched a student-driven predictive analytics challenge using INFORMS member survey data to uncover insights into stakeholder engagement, satisfaction, and participation. By identifying key drivers of retention and spotlighting underserved segments, the initiative generated actionable recommendations to strengthen mentorship, accessibility, and industry-academic collaboration. In parallel, a panel on algorithmic fairness invited dialogue on ethical model design and bias mitigation. Together, the challenge and discussion fostered community engagement while empowering the next generation of analytics talent to contribute to a more inclusive and data-informed INFORMS future.

6 - Sustaining Momentum: Advancing the INFORMS K-12 Education Outreach and Networking Program

Fenglian Pan, University of Arizona, Tucson, AZ, United States, Ken Murphy

This flash talk highlights the power of cross-society collaboration within INFORMS to tackle complex, real-world challenges through shared expertise in operations research, analytics, and data science. By bridging academic, industry, and educational communities—including outreach to K–12—we amplify our impact and inspire the next generation of problem-solvers. Whether optimizing supply chains or designing equitable public systems, our collective efforts demonstrate how interdisciplinary teamwork drives innovation and societal benefit.

7 - Understanding INFORMS Membership Dynamics and Trends from 2022-2024

Bing Si, Arizona State University, Tempe, AZ, United States, Alec Fishbach, Bo Li, Daniel Scott

Expanding participation amongst different member groups is essential to building a more inclusive professional community. In alignment with INFORMS' commitment to supporting a broad range of perspectives and backgrounds, this project proposes a data-driven initiative to analyze membership dynamics and trends from 2022 to 2024. Using survey data, the project seeks to uncover underlying organizational challenges and offer actionable strategies to improve member retention and reduce attrition across various member groups. By applying advanced analytics to INFORMS survey data, the study aims to generate meaningful insights that reinforce INFORMS' efforts to cultivate a collaborative, dynamic, and well-rounded professional community.

8 - Creating Pathways of Belonging in Supply Chain Management and Business Analytics

Afrooz Moatari-Kazerouni, Widener University, Chester, PA, United States, Jeff Lolli

Ensuring accessible and equitable career pathways in Supply Chain Management (SCM) and Business Analytics (BA) is critical for nurturing innovation and improving organizational performance. Nevertheless, existing literature highlights continual barriers for underrepresented students, including limited exposure to career opportunities, lack of mentorship, and a diminished sense of belonging within these fields. This study investigates how structured industry engagement can enhance students' perceptions, interest, and confidence in pursuing SCM and BA education and careers.

A cohort of high school students who identified as first-generation and/or students of color participated in a one-day industry event at an area venue, which included engaging activities such as tours, mentorship roundtables, career panels, and data analytics challenges. Using a web-based survey administered during the event, data were collected on the students' prior knowledge, perceived barriers, career interests, and confidence levels before and after participation. Quantitative analysis suggested that structured exposure programs not only increase awareness but also cultivate a stronger sense of belonging, which literature identifies as a key predictor of academic persistence and career attainment. This research offers evidence-based recommendations for designing outreach initiatives that effectively support underrepresented students in developing sustainable interest and engagement in SCM and BA, as well as ways to educate them on the vast and lucrative career paths available for post-secondary graduates in SCM and BA fields.

9 - Fostering Inclusive and Equitable Teaching and Learning Environments in Data Science and OR/MS Classrooms

Ying Lin, University of Houston, Houston, TX, United States, Yeawon Yoo, Woody Zhu

This project aimed to promote more inclusive and equitable teaching and learning environments in data science and OR/MS classrooms through three main activities. First, we hosted a webinar on culturally responsive instruction, where the guest speaker discussed core practices of culturally relevant pedagogy and how they can be meaningfully applied in data-centric courses. The session also addressed different entry points for implementing these approaches, common pitfalls to avoid, and strategies for self-reflection—emphasizing that culturally responsive teaching is an ongoing journey rather than a fixed goal. Second, we hosted the INFORMS Data Mining Society Teaching Award to recognize exemplary educators who applied innovative and inclusive strategies in data science and OR/MS instruction. Lastly, we conducted interviews with instructors and students with disabilities to better understand the barriers they face and identify practical solutions to improve accessibility and engagement in technical classrooms.

MD06

Building A Level 3 A306

Pivotal Role of Digital Innovation

Invited Session

eBusiness

Chair: Chenzhang Bao, Oklahoma State University, Tulsa, OK, United States

1 - The Effect of Generative AI-Generated Introductions on Early Fundraising Outcomes in Medical Crowdfunding

Wen Zhang, Baylor University, Waco, TX, United States, Yue Tang, Ruomeng Cui, Guihua Wang

In this study, we investigate the effect of using generative AI on early fundraising outcomes for medical crowdfunding campaigns. Using a dataset of medical-related campaigns from a leading crowdfunding platform, we analyze the effect of AI-generated campaign introductions on campaign donations in the first month after launch. Our results show that using AI-generated introductions can increase the total donation amount by 8.32% in the first month. Additionally, we find using AI-generated introductions can increase the total number of donations by 9.07% in the first month.

2 - Reducing No-Shows, Improving Continuity: The Combined Impact of Patient Interventions and Telehealth on Healthcare Utilization

Jiajia Qu, James Madison University, Harrisonburg, VA, United States

This study examines how patient intervention programs and telehealth services jointly influence healthcare appointment adherence and follow-up care patterns. Using patient medical history from a Northeastern U.S. healthcare system, we employ a robust analytical framework combining mixed-effects modeling to account for individual patient variability, difference-in-differences to assess program impacts, and propensity score matching to address selection bias. Our analysis evaluates both the

independent and synergistic effects of these interventions, with particular attention to their role in mitigating disparities across demographic and geographic groups. The findings provide critical insights for designing integrated care delivery systems that optimize access, continuity, and equity in healthcare utilization.

3 - The Effect of Platform Intervention Policies on AI Transparency: An Empirical Investigation of AI Labeling and User Engagement

Qinhui Wang, Oklahoma State University, Stillwater, OK, United States

The rapid proliferation of AI-generated content (AIGC) on digital platforms—particularly deepfakes involving synthetic images, audio, and videos—has intensified concerns about misinformation, content authenticity, and user trust. Social media users can now effortlessly generate and disseminate hyper-realistic synthetic media, which may obscure authentic content and facilitate malicious manipulation. In response, major platforms have implemented AI labeling policies to enhance transparency and mitigate the societal risks posed by deepfakes. This study empirically examines the effectiveness of such interventions by analyzing the relationship between AI label disclosure and user engagement, and how perceived video authenticity moderates this effect. Drawing on the Elaboration Likelihood Model (ELM), we theorize that AI labeling serves as a peripheral cue that negatively influences user engagement, particularly when content is flagged as synthetic. Moreover, we posit that higher levels of perceived authenticity may exacerbate this negative effect, as users become more sensitive to potential deception. To test these hypotheses, we analyzed a dataset of 809 TikTok videos containing 115,053 user comments, generated by 32 unique content creators between November 12, 2024, and April 14, 2025. The findings aim to inform policy and design strategies for platform governance in an era increasingly shaped by generative AI.

MD07

Building A Level 3 A307

Data-driven optimization

Job Market Showcase

Job Market Showcase

Chair: Yunlu Chen, Northwestern University, Evanston, IL, United States

1 - A Novel Methodology for Sensitivity Analysis of Combinatorial Optimization Using Predictive Modeling

Nishit Bhavsar, McMaster University , Hamilton, ON, Canada, Elkafi Hassini, Manish Verma

Mixed Integer Programming (MIP) is widely used for real-world applications such as resource allocation and network scheduling. However, its reliance on deterministic parameters limits its ability to generate long-term, reliable solutions. While stochastic and robust optimization address parameter variability, we propose a novel approach that integrates optimization and predictive analytics to enhance solution reliability. Our methodology involves three steps: (1) solving the MIP for multiple instances with varying deterministic parameters, (2) compiling optimization results into a database, and (3) training a predictive model to uncover patterns. We apply this to a MIP optimizing rail-truck intermodal networks for hazardous materials shipments, minimizing transportation costs and risks. The model determines intermodal routes, terminal selection, train connections, and drayage routes using realistic Canadian and U.S. networks. This approach improves MIP reliability by leveraging predictive insights, offering a more practical decision-making tool for complex transportation networks.

2 - Data-Driven Decision-Making in Healthcare Access Modeling: a Comparative Framework Analysis

Yujia Xie, Georgia Tech, Atlanta, GA, United States

Effective policymaking in healthcare access requires modeling that captures dynamics between healthcare supply and demand, further identifying opportunities to address barriers to receive adequate care. However, the uncertainty in the input data, the patient-provider interactions, the various system constraints, and the proper measurement for assessing access adequacy, all pose challenges in modeling and its use in decision making.

This presentation features two distinct modeling frameworks—simulation and optimization—applied to model healthcare access. Each demonstrates unique methodological challenges, spanning from data acquisition and input modeling to model design choices. The comparison between the two mathematical modeling approaches will highlight the different types of evaluations and the distinct policy insights each model can provide. For example, an optimization model can effectively optimize for a targeted performance measure, providing an optimal view of access, suitable for high-level (e.g., population level) insights. In contrast, a simulation model offers more granular statistics, providing a more realistic assessment of system behavior, especially under input uncertainty, however it lacks the efficiency of direct optimization for specific measures.

Acknowledging these unique input requirements, modeling strengths, and evaluation considerations is crucial for effective data-driven decision-making. This presentation illustrates the comparisons over a practical application focused on evaluating access to pediatric psychosocial services within a Medicaid-insured population.

3 - Scalable Linear Programming Based Optimization Algorithms for Data Science Applications

Yakun Wang, Lehigh University, Bethlehem, PA, United States

Many problems in data science—such as inference in graphical models and clustering—lead to mixed-integer nonlinear optimization problems that are computationally intractable. Existing methods often rely on heuristics with limited optimality guarantees. In this talk, I will present a line of research focused on developing strong linear programming (LP) relaxations that enable principled and scalable global optimization for these challenging problems.

In the first part, I address the inference problem in higher-order undirected graphical models. We reformulate the problem as a binary polynomial optimization and propose several LP relaxations. Applied to two applications—image denoising and error-correcting code decoding—our approach demonstrates that the LP relaxation is nearly always tight for image denoising and significantly outperforms existing LP-based methods in the decoding tasks.

In the second part, I focus on clustering. Building on a strong LP relaxation for the K-means clustering problem, we develop a scalable algorithmic framework for global optimization. This framework includes a separation algorithm for cutting-planes and leverages GPU-accelerated first-order solvers for LP to handle large-scale instances. To demonstrate the framework's flexibility, we incorporate several fairness notions from the literature for K-means clustering and extend the approach to spectral clustering. Extensive numerical experiments demonstrate that our framework can solve K-means problems to global optimality for datasets with up to 4,000 points, with promising applications in social network analysis and fairness-aware machine learning.

4 - On Feature Learning in Layered Models: An Optimization Perspective

Yunlu Chen, Northwestern University, Evanston, IL, United States

Layered models, such as neural networks, extract meaningful features through empirical risk minimization, yet the principles behind this process remain unclear. We analyze a two-layer nonparametric regression model akin to neural networks and prove that it naturally induces dimensionality reduction by identifying feature subspaces relevant for prediction—without explicit regularization. Our results explain this implicit regularization through set identifiability in optimization, showing how "sharpness" in the optimization landscape naturally enforces low-complexity solutions in finite samples.

MD08

Building A Level 3 A310

Technology Strategy for Managers and Policymakers

Invited Session

Information Systems

Chair: Naveen Thotadamane Basavaraj, Indiana University, Bargsersville, IN, United States

1 - Betting on Crime? the Impact of Legal spORts Gambling on Criminal Activity in the U.S

Kailing Mao, Carnegie Mellon University, Pittsburgh, PA, United States

The rapid expansion of legalized sports gambling has introduced significant societal problems that extend beyond individual financial consequences to broader social outcomes. Although previous research has linked sports betting to problem gambling, financial distress, and mental health concerns, its second-order effects remain underexplored. This paper examines the impact of widespread sports gambling on property crimes, particularly with the addition of an online channel. Furthermore, we also found young men are more likely to commit these crimes compared to their female counterparts, post-relative to pre-legalizing sports betting. The findings contribute to the growing discourse on the unintended social consequences of sports gambling and iGambling, emphasizing the need for regulatory measures to mitigate its broader societal impact.

2 - Clean up Your Act: the Regulation of Sexually Suggestive Live-Streaming Content

Yuan Yuan, Carnegie Mellon University, Pittsburgh, PA, United States

As live-streaming becomes increasingly prominent, how influencers present themselves to the public in real time has become central to online interactions. One controversial but unexplored behavior in this space is the use of private information, especially sexually suggestive content, to attract viewers. We leverage rich data from Douyu.com—a leading Chinese live-streaming platform, including audience activity logs, video content, and a platform-wide policy banning such content—to provide a comprehensive analysis of its prevalence, impact, and regulation. Using multi-modal video analytics that integrate computer vision and language models, we extract and quantify sexually suggestive content and find it to be positively associated with higher audience engagement, highlighting economic incentives for influencers. We causally identify the effects of the policy shock—the ban on sexually suggestive content on Douyu platform—and show that it caused significant declines in both audience engagement (demand) and content production (supply), highlighting the economic costs of content regulation. The decline is particularly pronounced in appearance focused content categories where influencer appearance is central to the content. We develop a stylized game-theoretic model that surprisingly predicts that the impact of policy shock may not be universally negative. The policy reduces the supply of cheap (to produce) sexually suggestive content and causes some influencers to exit, thus lowering competition. This can incentivize a segment of influencers to invest more in content creation. Thus, the negative impact of content regulation can be offset for this segment of influencers.

3 - Information Integration and Hospital Performance: the Role of Managerial Ability

Tianjian Guo, Indiana University, Indianapolis, IN, United States

Alternative payment models (APMs) have emerged as a key strategy for improving care quality while ensuring financial sustainability in the U.S. This paper examines how two organizational capabilities—managerial ability (MA) and information integration capability (IIC)—jointly influence hospital APM adoption and operating margin. Grounded in the resource-based view and dynamic capability theory, we hypothesize that hospital-level MA, IIC, and their interaction positively affect both outcomes. Using a panel dataset of over 18,000 hospital-year observations from 1,904 U.S. hospitals, we find that MA and IIC each have a positive effect on operating margins, and that their interaction has a complementary effect on both outcomes. Further analysis reveals that MA and IIC have distinct effects on the adoption of specific APM types with different levels of financial risk. These findings offer important policy and managerial implications, highlighting how hospitals can strategically leverage managerial and IT capabilities to adapt to the evolving healthcare landscape.

MD09

Building A Level 3 A311

Human-algorithm Interactions

Job Market Showcase

Job Market Showcase

Chair: Qiaowen Guo, Washington University in St. Louis, St. Louis, 63130, United States

1 - Whose Ideas Advance? Power, Status, and Decision-Making in Human and AI Design Teams

Yakira Mirabito, Massachusetts Institute of Technology, Cambridge, MA, United States

Design decisions in engineering are shaped not only by technical considerations, but also by complex organizational dynamics. My research investigates how hierarchy, including power and status, affects the evaluation and selection of design ideas, and how these dynamics evolve as AI becomes more integrated into engineering work. This talk draws from three studies that investigate how team dynamics and AI tools influence decision-making in engineering contexts. First, an observational study of industry design reviews highlights how hierarchy affects whose feedback is considered. Building on these insights, I conducted controlled experiments that manipulated participants' sense of power and status to examine how these factors shape design evaluations. A follow-up study extends this work by exploring how power dynamics affect the evaluation and selection of AI-generated versus human-generated design concepts. In this talk, I will briefly introduce the first two studies and focus on the third, which examines how power influences the evaluation and selection of AI-generated ideas. Together, these studies reveal how hierarchy and emerging technologies interact to influence design outcomes.

2 - Improving Human-AI Workflows for Clinical Decision-Making

Sarah Jabbour, University of Michigan, Ann Arbor, MI, United States

Accurate diagnosis is critical to the care of hospitalized patients, yet diagnostic errors remain common. While AI tools have the potential to support clinicians in making complex decisions, their real-world impact depends on how clinicians interact with them. Even high-performing models can mislead clinicians when deployed, especially in the presence of automation bias—where humans over-rely on AI even when it is wrong. Most models are optimized in isolation, without regard for how humans incorporate predictions into their decision-making. To address this, we conducted two large-scale randomized clinical vignette studies evaluating two automation bias mitigation strategies: instance-based image explanations and selective prediction. We found that image explanations do not help clinicians recover from inaccurate AI, while selective prediction does reduce automation bias—but at the cost of increased false negatives, potentially leading to missed diagnoses. These findings highlight the need for human-centered evaluation frameworks that explicitly account for clinician-AI interaction.

3 - Accuracy Obsession: Humans Prioritize Immaterial AI Accuracy over Their Own Compensation Unless We Educate Them

Matthew DosSantos DiSorbo, Harvard Business School, Boston, MA, United States, Kris Johnson Ferreira

Problem Definition: Accuracy isn't everything. Often high or low sensitivity is more important than overall correctness — however, we find that humans prefer accurate algorithms, even to their detriment. We identify and address an accuracy obsession: humans trust algorithms based on the algorithm's accuracy, not the actual monetary reward they could earn by using the algorithm.

Methodology/results: We model a two-stage algorithm that estimates the probability of a positive case and, if that probability exceeds a pre-determined threshold, recommends a human review; the human then makes the final decision. Penalties and rewards incurred by the human vary across outcomes. However, we introduce theory suggesting humans have an accuracy obsession: trust in the algorithm's recommendations is determined by algorithm accuracy, not performance (reward earned by following the algorithm's recommendations). Online lab experiments confirm our hypotheses: participants trusted a sub-optimal algorithm calibrated to maximize accuracy, not reward, much more than an optimal algorithm calibrated to maximize reward. These results held across two contrasting settings: high and low penalty for false negatives and, in turn, an optimal algorithm that is much more, and much less, sensitive than the sub-optimal algorithm. Finally, we design an educational intervention that mitigates the accuracy obsession.

Managerial Implications: We demonstrate that humans are biased towards accuracy obsession: trusting algorithms based on their accuracy, not their expected reward. We introduce a simple educational intervention that mitigates this bias, and hope our work will help managers prepare their employees for success in the many human-AI collaborations where accuracy isn't everything.

4 - The AI Management Paradox: How Algorithmic Control Systems Drive Efficiency but Erode Consumer Trust

Qiaowen Guo, Washington University in St. Louis, St. Louis, MO, United States, Xiang Hui, Fuqiang Zhang, TIANJUN FENG

<p break-words="">Can AI management systems optimize workforce performance without sacrificing the human elements that build consumer trust? I investigate this paradox through a large-scale field experiment at a leading e-commerce platform, randomly assigning 265,967 customer service interactions to either algorithmically-managed gig workers or traditionally-supervised employees.

<p break-words="">The results reveal a stark efficiency-trust tradeoff. Algorithmic control excels at driving measurable performance: gig workers achieve 35% faster service, 4% higher case closure rates, and 30% quicker response times. However, these efficiency gains come at a profound cost to consumer trust. Customers served by algorithmically-managed workers express 7% more negative emotions and subsequently reduce their platform purchases.

<p break-words="">Using machine learning text analysis, I uncover the mechanism driving this paradox: AI's narrow metric focus creates exploitable loopholes that workers strategically game. Nearly half (47.4%) of gig workers' case transfers are unnecessary escalations designed to boost performance and compensation. These strategic behaviors don't just frustrate consumers—they fundamentally erode trust. Following strategic transfers, consumers shift purchases from high-value durable goods to low-cost consumables, a 9.65% reduction in average product price that signals diminished platform confidence.

<p break-words="">This research exposes a critical vulnerability in AI management systems: optimizing for measurable efficiency can systematically degrade unmeasurable trust. As millions of workers globally transition to algorithmic management, my findings provide essential insights for designing AI systems that balance operational efficiency with the human elements necessary for sustained consumer relationships. The future of workforce management depends on resolving this paradox.

MD10

Building A Level 3 A312

Machine Learning and Optimization

Job Market Showcase

Job Market Showcase

Chair: Nathan Justin, University of Southern California, Los Angeles, CA, United States

1 - Interplay of Global Optimization and Machine Learning

Pariyakorn Maneekul, University of Washington, Seattle, WA, United States

High-dimensional black-box optimization is a prevalent challenge in science and engineering. Through the interplay between optimization and machine learning, we develop adaptive search algorithms that balance exploration, exploitation, and estimation by leveraging machine learning to construct efficient surrogate models. Three primary contributions are presented. First, a multi-level Partitioning and Branch-and-Bound (PBnB) algorithm is proposed for level-set approximation. This method enhances the original PBnB by using importance sampling to identify promising subregions and integrating Gaussian processes to guide local exploration, with theoretical bounds on the approximation quality.

Second, to address high dimensionality, the Branching Adaptive Surrogate Search Optimization (BASSO) framework is introduced. A finite-time analysis proves that BASSO can achieve performance that scales linearly with dimension under specific theoretical assumptions. Implementations using Gaussian process and regularized quadratic regression surrogates are explored to balance local exploitation with global exploration, and numerical experiments investigate the gap between theory and practice.

Finally, for noisy functions, the dissertation extends the Single Observation Search Algorithm (SOSA). It replaces neighborhood averaging with a more robust quadratic regression for function estimation and incorporates optimistic sampling from reinforcement learning to improve the search strategy. Theoretical convergence results for this enhanced SOSA are provided, and numerical experiments demonstrate significant performance gains over the baseline algorithm. This research culminates in a discussion of future applications, including the potential of quantum computing for global optimization

2 - Learning Optimal Prescriptive Trees Robust to Distribution Shifts

Nathan Justin, University of Southern California, Los Angeles, CA, United States, Daniël Vos, Phebe Vayanos

In domains such as personalized medicine, historical data is used to learn what treatments to prescribe to maximize positive outcomes. Previous studies have proposed methods for creating prescriptive trees: human-interpretable diagrams that indicate what type of treatment an individual should get based on their measurements. However, a remaining problem is that the models perform worse over time when there is a shift in the data collection process or when data from a different source is used during the training and prediction phases. To solve this problem, we propose a method that considers data uncertainty by optimizing distributionally robust prescriptive trees. We formulate a linear-time algorithm to find the worst-case distribution shift within a given Wasserstein distance around the dataset and use it as a subroutine within the main problem. Our algorithm does not depend on any specific causal effect estimator and can, therefore, be applied in various contexts.

3 - Scalable First-Order Method for Certifying Optimality in Interpretable Machine Learning

Jiachang Liu, Cornell University, Ithaca, NY, United States, Soroosh Shafiee, Andrea Lodi

For interpretable machine learning in high-stakes decision-making and scientific knowledge discovery, ℓ_0 optimization has emerged as a powerful technique. Unlike past approximation-based methods, it can yield extremely sparse models without compromising accuracy --- even in the presence of high-dimensional and highly correlated features. However, solving the ℓ_0 problem to optimality remains computationally challenging for large-scale datasets. A major bottleneck lies in the expensive computation of dual bounds in the branch-and-bound framework.

To address this, we develop a first-order proximal gradient method to solve the perspective relaxation of the original cardinality-constrained problem. We reformulate the relaxation as a composite optimization problem and show that the proximal operator of its non-smooth component can be computed exactly in log-linear time complexity, eliminating the need to solve an expensive second-order cone program. We further introduce simple restart strategies that achieve the optimal linear convergence rate while preserving low per-iteration computational cost. Lastly, we design our algorithm with GPU-compatibility in mind, as matrix-vector multiplication is the most computationally expensive step; this allows us to significantly accelerate our method using GPUs.

Extensive experiments on synthetic and real-world datasets demonstrate that our method significantly accelerates dual bound computations and is effective in providing optimality certificates for large-scale problems. Such a success opens doors to new research opportunities in developing GPU-friendly first-order methods in both machine learning and mixed-integer nonlinear programming.

MD11

Building A Level 3 A313

High-dimensional Data Analysis II

Job Market Showcase

Job Market Showcase

Chair: Wuxia Chen, University of Pittsburgh, Pittsburgh, PA, 15224-2136, United States

1 - Tensor-Based Predictive Modeling and Control for High-Dimensional Streaming Data

Zihan Zhang, Georgia Institute of Technology, Atlanta, GA, United States

The rapid advancement of sensing technologies has led to an explosion of high-dimensional data—from high-resolution images and videos to 3D point clouds—that fundamentally challenge traditional control methodologies. Conventional approaches, rooted in low-dimensional signal processing, fail to capture the complex spatio-

temporal dependencies inherent in such data. Simple vectorization disrupts essential structural information, while many learning-based methods require extensive datasets and lack rigorous theoretical foundations.

This talk presents a suite of tensor-based process control frameworks that preserve spatio-temporal structure while enabling principled control analysis. I will introduce strategies for incomplete sensing scenarios, novel system modeling approaches that capture localized correlations, and dynamic controller optimization methods. The framework naturally extends to address nonlinear spatio-temporal correlations via diffusion models, while also providing uncertainty quantification capabilities.

These contributions establish a data-efficient, theoretically grounded paradigm for controlling modern intelligent systems operating with high-dimensional, multimodal sensory data.

2 - Adaptive Tensor and Multimodal Learning for Human-Centered Systems

Elif Konyar, Georgia Institute of Technology, Atlanta, GA, United States, Kamran Paynabar

Modern engineering and healthcare systems generate high-dimensional, multimodal, and distributed data that capture complex interactions between humans, machines, and environments. Leveraging these data requires methods that are privacy-preserving, interpretable, and robust to heterogeneity and irregularity. In this talk, adaptive tensor and multimodal learning methods will be presented that advance two complementary directions: human-centered systems, which use interpretable and privacy-preserving models to enhance decision-making in healthcare, manufacturing, and transportation; and human-in-the-loop systems, which directly incorporate human behaviors and expertise to enable adaptive, user-aware AI. The talk will cover federated tensor regression for decentralized telemonitoring and industrial fault detection, semi-supervised tensor decomposition for computational phenotyping from electronic health records, and robust tensor models for anomaly detection in manufacturing and transportation networks. It will also highlight multimodal modeling of gaze, head, and hand motion in augmented reality-guided tasks. Together, these contributions demonstrate how tensor and multimodal learning can support trustworthy, personalized, and adaptive human-AI systems.

3 - Directed Graphical Models for Root Cause Analysis in Multimodal Systems Collecting Heterogeneous Data

Manni Zhang, Purdue University, West Lafayette, IN, United States, Ana Estrada Gómez

Nowadays, complex systems are equipped with multiple sensors that collect high-dimensional and heterogeneous data, including scalars, functional signals or profiles, and images. Such systems are prone to failures, and diagnosis is hard as the intricate interactions among the variables mask the root causes for fault diagnosis. Directed graphical models (DGMs) are a powerful tool for representing the probabilistic relationships between variables in such systems. Nonetheless, frameworks learning DGMs with mixed types of data are scarce. Existing methods focus on either scalar variables or profile data, or have stringent assumptions. In this article, we propose a framework, tensorDGM, for all data types and learn the structure of the system for root-cause analysis. We learn the structure of the DGM by fitting penalized tensor-on-tensor regressions with group Lasso penalties for variable selection and an L2 penalty to handle the group selection of variables. To solve the optimization problem, we propose a cyclic coordinate accelerated proximal gradient descent algorithm with Bayesian optimization for hyperparameter tuning. Through simulations and a case study, we illustrate the advantages of the proposed method, tensorDGM.

MD12

Building A Level 3 A314

Managing Disruptions

Job Market Showcase

Job Market Showcase

Chair: Ozoemenam Ohagwu, Walden University, 100 Washington Square, Minneapolis, MN, 55401, United States

1 - The Role of Firm Risk in Supply Chain Network Adaptation

Shailesh Divey, MIT, Cambridge, MA, United States

Problem Definition: This paper investigates the impact of a firm's financial risk on its position within a supply chain network. Given the increasing complexity of supply chains and the role of financial stability in shaping business partnerships, understanding these dynamics is crucial for both firms and policymakers. **Methodology/Results:** We analyze supplier-customer relationship data from the US manufacturing industry using network analysis techniques. Firm risk is measured through idiosyncratic volatility derived from the Fama-French three-factor model with industry effects asset pricing model, and supply chain centrality is captured using degree centrality and eigenvector centrality. Our study employs fixed effects regression models and dynamic panel estimation techniques to establish relationships between firm risk and network position. We show how a firm's centrality adapts to a change in its risk exposure. **Managerial Implications:** This study provides critical insights for managers and policymakers aiming to strengthen supply chain resilience. Firms must proactively manage financial risk to maintain strategic supplier and customer relationships. Monitoring network structures and implementing strategies like supplier diversification and credit risk management can mitigate adverse effects. Additionally, high-risk firms can explore contractual mechanisms or risk-sharing agreements to sustain strong supply chain ties. Academics studying supply chain adaptation and risk propagation can build upon these findings to further explore the intersection of financial stability and network resilience.

2 - Managing Supply Disruption Risk: Selecting Unreliable Suppliers Under Cardinality Constraints

Ning Ma, Arizona State University, Tempe, AZ, United States, Heng Zhang, Yimin Wang

Recent natural disasters have significantly heightened the risk of supply disruptions, underscoring the urgent need to improve our understanding of how to effectively configure supply bases to manage unreliable supplier yields. While the literature has extensively examined the trade-off between cost and reliability, it has largely overlooked constraints on the supplier base—particularly cardinality constraints that naturally reflects considerations of cost control, operational efficiency, and the development of stronger supplier relationships.

This study addresses this gap and enhances the existing body of knowledge regarding the optimal selection of suppliers within a supply base. We demonstrate that the prevailing approach of ranking suppliers based on cost and reliability is not universally applicable when the cardinality constraint is introduced. Moreover, we show that the supply base design problem does not exhibit desirable mathematical properties, such as submodularity, which presents significant challenges for its analytical treatment.

To overcome these challenges, we propose a uniform demand approximation (UDA) schedule, which promotes a surrogate optimization problem that approximates the original problem. We derive a fully polynomial-time approximation scheme (FPTAS) for this problem. Additionally, we prove that the surrogate problem exhibits the ordered-submodularity property, enabling the use of efficient greedy-type algorithms when the supplier base is subject to cardinality, knapsack, or matroid constraints. Through numerical studies, we demonstrate that our solution is highly efficient and outperforms all benchmarks. Furthermore, we provide managerial insights into the structure of the optimal supply base configuration.

3 - On the Pursuit of ESG and Firm Resilience

WENTING LI, Arizona State University, Tempe, AZ, United States

Problem Definition: Environmental, social, and governance (ESG) practices, while broadly adopted, are increasingly questioned for their value. Amid ongoing global disruptions, executives question whether ESG is a discretionary cost that harms firm performance or a source of strategic value. While extensive research exists on ESG and firm performance, empirical evidence disentangling its impact during crises, particularly within the supply chain context, remains scarce. **Methodology/Results:** Leveraging the COVID-19 pandemic as a natural experiment, we assess firms' competitiveness relative to industry peers amid crises. Our instrumental variable and matching estimations reveal that while ESG adoption has no significant effect on firm performance under stable market conditions, it becomes a key differentiator in crises. A standard deviation increase in ESG score leads to a 0.107 unit (1.9%) rise in excess return on assets (ROA) compared to industry peers. Moreover, our

mechanism analyses uncover two distinct pathways of this ESG-enabled competitiveness: (1) supplier-driven cost efficiencies and (2) buyer-driven ESG price premiums. We show that the cost efficiency pathway stands out as the primary source of competitive advantage during disruptions. In comparison, the ESG-driven price premiums pathway may be present during normal times but fades during disruptions. Managerial implications: Our study suggests that ESG adoption may not benefit firms during normal times but becomes crucial during disruptions when resources are scarce. With heightened global disruptions, our study offers compelling evidence for a paradigm shift in ESG investment from social responsibility to supply chain resilience.

4 - Effective Strategies to Mitigate the Impact of Cargo Shipment Abandonment

Ozoemenam Ohagwu, Walden University , Minneapolis, MN, United States

Business leaders in multiple U.S. exporters refuse to pay freight charges and abandon their cargo shipments, leaving shipping business leaders with monetary losses. Some small and medium-sized U.S.-based shippers lack effective strategies to address the risk of cargo shipment abandonments. Grounded in the theory of constraints, the purpose of this qualitative pragmatic inquiry study was to explore the effective strategies used by small and medium-sized U.S.-based shippers to successfully mitigate the risk of cargo shipment abandonment. The participants included three U.S. shipping industry leaders from non-vessel-operating common carriers and three from vessel-operating common carriers, all with an average of 25 years of experience and a demonstrated ability to effectively mitigate cargo shipment abandonment risks. Data were gathered through semistructured interviews and publicly accessible information. Through thematic analysis, eight key themes emerged: relationship management, disruptions, financial management, risk management and recovery, service quality, technology, unclaimed and abandoned cargo, and service category. A key recommendation is for U.S.-based shippers to employ effective communication and collaboration with stakeholders. The implications for positive social change include the potential for U.S.-based shippers to effectively manage maritime supply chain disruptions. This capability reduces cargo shipment abandonment and operational costs, leading to lower local and international commodity prices. Ultimately, it enhances community sustainability, creates new job opportunities, and ensures tax revenues that can bolster local financial stability.

MD13

Building A Level 3 A315

Industry Job Search Panel

Panel Session

Career Fair

Co-Chair: Beverly Wright, Wavicle Data Solutions, Anderson, SC, United States

1 - Moderator Panelist

Beverly Wright, Wavicle Data Solutions, Anderson, SC, United States

The purpose of this session is to bring visibility to the students and postdocs looking for non-academic positions. Panelists from business and industry will share their experiences. This panel discusses their experiences in the interview process along with the do's & don'ts associated with the job search.

2 - Panelist

Wes Chaar, Avant-Garde Strategy, Dallas, TX, United States

3 - Panelist

Samaneh Shiri, UPS, Timonium, MD, United States

4 - Panelist

Abdul Rahman Noordeen, Ericsson, Cummings, GA, United States

5 - Panelist

MD14

Building A Level 3 A316

Multi-Objective Optimization

Contributed Session

Contributed

Chair: Dhiraj Pokhrel, University of Tennessee, Knoxville, Knoxville, TN, United States

1 - Preference Optimization on Pareto Sets: on a Theory of Multi-Objective Optimization

Abhishek Roy, Texas A&M University, College Station, TX, United States, Geelon So, Yi-An Ma

In multi-objective optimization, a single decision vector must balance the trade-offs between many objectives. Solutions achieving an optimal trade-off are said to be Pareto optimal---these are decision vectors for which improving any one objective must come at a cost to another. Many decisions can be Pareto optimal, so when the decision maker can choose only one, this raises questions of which solution to pick and how. We formulate the problem as $\text{Pareto-constrained optimization}$, where the goal is to optimize a preference function constrained to a set of Pareto optimal solutions. This constrained optimization problem poses significant challenges: not only is the constraint set defined implicitly, but it is also generally non-convex and non-smooth, even when the objectives are strongly convex. We propose a reformulation of the problem where the constraint set is redefined in terms of an appropriately defined manifold. This reformulation allows us to introduce a clearer and more accurate notion of optimality and stationarity, improving upon existing definitions in the literature. We provide an algorithm with a last-iterate convergence rate of $O(K^{-1/2})$ to this notion of stationarity when the preference function is Lipschitz smooth and when the objective functions are strongly convex and Lipschitz smooth. Moreover, motivated by practical applications like Reinforcement Learning with Human Feedback (RLHF), we extend this algorithm to tackle the case where the preference function can only be queried via dueling feedback.

2 - On Presaging Multiple-Objective Capital Asset Pricing Models by Multiple-Objective Portfolio Selection: Originating Multiple Risk-Free Assets and Proving the Tangency to Expected Utility Indifference Surfaces

Yue Qi, Nankai University, Tianjin, China, People's Republic of, ZhuJun Qi, Yue Wang

Markowitz originates portfolio selection as two-objective optimization. Sharpe considers one risk-free asset for portfolio selection and originates his capital asset pricing model as one center-piece of finance. Researchers contemplate capital allocation as constructing portfolios by one risky asset and one risk-free asset. Researchers maximize expected utility by locating the tangent point between a capital allocation line and an expected utility indifference curve. After their feat, Markowitz and Sharpe notice extra objectives. Researchers are advancing multiple-objective portfolio selection. However, there are still limitations in leaping from multiple-objective portfolio selection to multiple-objective capital asset pricing models. In such an area, this paper theoretically contributes to the literature. We originate multiple risk-free assets, extend traditional capital allocation, and prove the extension's plane property. We maximize expected utility by locating the tangency between a capital allocation plane and an expected utility indifference surface. Moreover, we prove that the tangency actually happens on a line instead of the traditional point. Overall, this paper theoretically acts as a foothold for eventually achieving multiple-objective capital asset pricing models.

3 - Pareto Set Domination Under Limited Information

Pubudu Jayasekara, The University of Tennessee at Chattanooga, Chattanooga, TN, United States, Margaret Wiecek

Mixed-integer multiobjective optimization problems typically yield nonconvex and disconnected sets of nondominated solutions, making the identification and analysis of the complete Pareto set particularly challenging. Their structural complexity demands the development of novel optimization approaches. While recent research has

increasingly focused on solving mixed-integer multiobjective problems, the concept of Pareto set domination, especially under limited information, has received relatively little attention.

This study addresses that gap by proposing a method to assess domination relationships among Pareto sets when complete information is unavailable. The proposed approach offers a structured way to compare partial or incomplete solution sets, which is particularly useful in iterative or heuristic-based optimization methods. Even with exact algorithms, generating the full Pareto set can be computationally expensive, making this approach broadly applicable. By evaluating dominance using only known nondominated points or subsets, the method enhances decision-making, and guides search strategies in complex problem settings.

Numerical examples are included to highlight its practicality and potential for wider use.

4 - Collision Avoidance and Path Finding in a Robotic Mobile Fulfillment System Using Multi-Objective Meta-Heuristics

Ahmad Kokhahi, Clemson university, Central, SC, United States, Mary Beth Kurz

This paper addresses the challenges of collision avoidance and task assignment in Robotic Mobile Fulfillment Systems (RMFS), where Automated Guided Vehicles (AGVs) coordinate to execute warehouse operations. While prior research primarily focuses on minimizing collisions and travel time, this study uniquely incorporates energy consumption into the optimization objectives. The proposed framework introduces a new rule to handle collisions by focusing on saving energy, based on whether AGVs are carrying a load or not. It also improves path planning using a modified A* algorithm that looks at both location and time, and changes routes in real time to avoid collisions, helping reduce delays and energy use.

For task assignment, two heuristic multi-objective approaches are developed: Non-Dominated Sorting Genetic Algorithm II (NSGA-II) and Adaptive Large Neighborhood Search (ALNS). Both techniques leverage custom neighborhood operators—swap, insertion, and arrangement—to cluster tasks by aisle proximity, minimizing travel distance and the likelihood of collisions. The NSGA-II algorithm incorporates a new Arithmetic Crossover (ARC) operator to better balance task distribution among AGVs.

Comprehensive computational experiments demonstrate the superior performance of the proposed methods in terms of both hypervolume and CSS metrics when compared to benchmark strategies. Notably, the improved NSGA-II algorithm consistently outperforms other variants. The findings suggest that energy-aware collision resolution and task assignment strategies can significantly enhance RMFS efficiency, offering valuable insights for warehouse operations management.

5 - Multi-Period Bioenergy Supply Chain Planning Incorporating Biomass Pricing

Leyla Tavassoli, Southern Methodist University, Dallas, TX, United States, Halit Uster

This study develops a planning framework for bioenergy supply chain networks that integrates both tactical and strategic decision-making by addressing spatial and temporal characteristics of the network design. We formulate a mixed-integer programming framework to capture the decisions on facility types and locations, their capacities, and multi-period shipment planning while taking into account farmers' biomass planting decisions. We implement decomposition-based algorithms and introduce enhancements to improve computational performance of our solution approach. The models are validated through computational experiments and a real-world case study in the Midwest United States using GIS-based spatial data.

6 - Optimizing Invasive Species Management with Bi-Objective Modeling and Gis Integration

Damitha Bandara, University of Tennessee-Chattanooga, Chattanooga, TN, United States, Lakmali Weerasena, Nyssa Hunt, Mark McKnight

Invasive species are costly and challenging to manage, highlighting the need for efficient control strategies. This study introduces a bi-objective optimization model integrated with GIS to support invasive species management (ISM). The model accounts for functional connectivity among infested sites and identifies optimal control strategies under varying minimum control levels. Applied to Chinese privet management across 317 acres in Chattanooga, Tennessee, the approach demonstrates how decision-makers' preferences and control targets influence outcomes. The results underscore the value of mathematical modeling in guiding ISM decisions and show how geographic visualization enhances practical strategy development.

7 - Optimizing Resource Allocation for Distribution System Upgrades for Resiliency Under Budget Constraints

Dhiraj Pokhrel, University of Tennessee, Knoxville, Knoxville, TN, United States, Hugh Medal

Severe blackout events due to extreme weather have been increasing in recent years, highlighting the need for resilient power distribution networks. In this research, we study the capacity expansion problem for upgrading the distribution grid under a budget constraint, with a focus on enhancing system resilience. We formulate our problem as a two-stage stochastic mixed-integer programming model to identify optimal investment decisions involving the placement and sizing of distributed generation units both renewable and non-renewable, as well as the placement of switches and line hardening measures. First stage decisions consist of infrastructure upgrades, while the second stage consists of operational responses post-disruption, aiming to minimize expected costs associated with generation, maintenance, and load shedding across all scenarios. We validate our model using the IEEE 33-bus test system, demonstrating the computational efficiency of our approach and the effectiveness of the proposed strategy in improving distribution grid resilience.

MD15

Building A Level 4 A401

Advancements in Airline Revenue Management and Pricing

Invited Session

Air Transportation Section

Chair: Yuxuan Lu, 01801

1 - Capturing Choice Behavior in Revenue Management Systems Through a Passenger Schedule Preference Model

Caroline Muhith, Georgia Institute of Technology, Atlanta, GA, United States, Laurie Garrow

Passenger schedule preference models inform passenger behavior and decision-making, recognizing that passengers value itinerary attributes in addition to price. Business passengers, for example, often prefer specific departure and arrival times or nonstop flights, reflecting a lower tolerance for schedule deviations. These preferences contribute to the perceived disutility of a flight and indirectly impact demand. Currently, many revenue management (RM) systems lack customer schedule preferences. This paper introduces a schedule preference model into the customer decision making process that integrates three utility components: (1) a nonlinear replanning disutility, which penalizes deviations from ideal departure and arrival windows using a sigmoid function with modifiable beta, slope, and offset parameters to reflect behavioral sensitivities across passenger segments; (2) unfavored airline disutility, derived from the total airline seat share at the originating airport, to capture frequent flyer effects; and (3) a path-quality index disutility that penalizes itineraries with connections. This model was implemented in PassengerSim, a new competitive simulation tool for RM applications developed by Alan Walker, Jeff Newman, and Laurie Garrow on a network comprising four airlines, 44 cities, and 10 different fare classes. Preliminary results show that the airline with the most favorable schedule characteristics—more departures during peak time and higher seat shares at the originating airport—saw revenue increases of 1.3% to 2.0%, and system load factor improvements of 0.9% to 1.4%. These findings show how incorporating a schedule preference model into RM systems can accurately forecast and capture how itinerary features and demand impact revenue.

2 - Estimation of Price Sensitivity Using Passengersim Choice Set Data

Antonio Ramirez, Georgia Institute of Technology, Atlanta, GA, United States

Airline pricing and revenue management analysts continue to seek to understand how passengers choose what to book when presented with a set of travel itinerary options that vary in characteristics such as airline, price, and fare class. Without recreating a complete and comprehensive theoretical model, this research aims to determine the

feasibility of using basic multinomial logistic regression (MNL) models, which implicitly capture only limited aspects of customer choice, to estimate passenger price sensitivity. The choice set data used in our MNL models come from PassengerSim, a proprietary airline revenue management simulation tool, and are a random sample of nearly 10,000 choice sets from a simulation of a network comprising of four airlines, 44 cities, and 10 fare classes. We interpret the price-related coefficients from the fitted MNL models and roughly-converted FRAT5 values – the ratio between fares at which 50% of passengers sell-up – to examine price sensitivities separately by passenger type and booking window. Our results suggest that MNL models that account for airline, nonstop flights, and fare ratio (defined as price of alternative scaled by lowest available price in the choice set) and that rely on purchases-only choice set data (i.e., excluding no-go passengers) may produce reasonable price sensitivity estimates. Although, models that include no-purchase alternatives appear to be more robust at producing valid price sensitivity estimates. These findings suggest that an airline could use its own historical choice set and booking decision data to estimate price sensitivities and better gauge customer willingness-to-pay.

3 - Improving Airline Demand Forecasting Using Macroeconomic Trends and Sentiment Analysis

Soheil Sibdari, University of Massachusetts Dartmouth, North Dartmouth, MA, United States

Enhancing precision in demand forecasting is essential for airline profitability and operational efficiency. This research integrates historical sales data with macroeconomic indicators—such as GDP growth, exchange rates, energy prices, and middle-class expansion—and consumer sentiment analysis to improve airline demand forecasts. By combining economic forecasts with sentiment signals from news and public data, the model captures both financial trends and shifts in traveler behavior. A mathematical framework and empirical validation show that this approach improves predictive accuracy and supports dynamic decision-making in pricing, route planning, and capacity allocation. Using time series and spatial analysis, the study also captures regional and temporal variation, offering airlines more responsive and location-specific demand forecasts.

4 - Revenue and Load Factor Trade-off in Revenue Management

Darius Walczak, PROS Inc., Houston, TX, United States

The primary objective of revenue management (RM) is to maximize expected revenue and sophisticated software systems have been deployed to achieve it. In practice though, users of RM systems need to know what expected bookings and thus load factors are achievable under those controls. These key business metrics rival revenue in importance but have historically received less attention from the modeling perspective in RM.

We focus on the single-leg dynamic setting for which there are results available to generate the so-called efficient frontier between revenue and expected bookings/load factor. The efficient frontier makes the tradeoff between the two key RM metrics precise and interpretable.

Then, we extend our analysis to RM models with price-sensitive demand by broadening the concept of the fare class efficiency ('concave envelope') and the associated transformed fares/fare adjustments to account for the load factor objective. We also suggest ways to map network level load factor objectives to the corresponding individual flight-leg level objectives.

5 - Managing Congestion in Advanced Air Mobility Operations Using a Neural Network Enhanced Bi-Level Optimization Approach

Haochen Wu, University of Michigan, Ann Arbor, Ann Arbor, MI, United States

Advanced Air Mobility (AAM) represents an evolution of the air transportation system by introducing low-altitude, potentially high-traffic environments. AAM operations will be enabled by both new aircraft, as well as new safety- and efficiency-critical supporting infrastructure. Published concepts of operations from both public and private sector entities establish notions such as federated management of the airspace and public-private partnerships for AAM air traffic, but there is a gap in the literature in terms of integrated tools that consider all three critical elements: AAM fleet operators (lower layer), airspace service providers (middle layer), and overall system governance from the legacy air navigation service provider (upper layer). In this work, we explore modeling congestion management within the AAM setting using a bi-level optimization approach, focusing on (1) time-varying, stochastic AAM demand, (2) differing congestion management strategies, and (3) the impact of unscheduled, "pop-up" demand. We show that our bi-level formulation can be tractably solved using a Neural Network-based surrogate which returns solution qualities within 0.1–5.2% of the optimal solution. Additionally, we show that our congestion management strategies can reduce congestion by 25.7–39.8% when compared to the scenario of no strategies being applied. Finally, we also show that while pop-up demand degrades congestion conditions, our congestion management strategies fare better against pop-up demand than the no strategy scenario. The work herein contributes a rigorous modeling and simulation tool to help evaluate future AAM traffic management concepts and strategies.

MD16

Building A Level 4 A402

Practical Applications

Invited Session

The Practice Section of Inform

Chair: Srinivas Bollapragada, GE Vernova Advanced Research Center, Niskayuna, NY, United States

1 - Real-Time Aircraft Engine-Informed Contrail Prediction and Joint Route and Airspace Optimization to Mitigate Co2 and Non-Co2 Aviation Emissions

Banu Gemici-Ozkan, Ph.D., GE Aerospace, Niskayuna, NY, United States, Saikat Ray Majumder, Guang Zhao, Anindya Bhaduri, Brett Matthews, Christopher Jerrett, Zhihui Yang, Jerrold Cline, George Dalakos

The aviation industry contributes to global warming through both CO2 emissions and non-CO2 effects. While the warming impact of CO2 emissions is well-established, accounting for approximately 2.5% of global anthropogenic contributions, non-CO2 effects—such as persistent contrails and NOx emissions—exert nearly double the radiative forcing compared to CO2 alone (Lee et al., 2021). In recent years, governments, private entities, and international organizations have intensified research efforts to quantify and validate the warming effects of contrails.

This presentation will summarize research conducted by GE Aerospace Research's Climate Sciences & Optimization Lab to minimize the overall warming effects of CO2 and non-CO2 emissions. The presentation will focus on two core areas:

Novel hybrid AI/ML and computer vision-based validation methods developed under the DOE ARPA-E-funded program, which introduces the Engine-Informed Prediction of Aviation-Induced Cirrus Trails (EPICTrails) system. EPICTrails integrates a real-time, in-flight prediction model to detect persistent contrails, validated at scale using satellite-based computer vision methodologies (achieving an F1-score ≥ 0.8 for contrails lasting 3+ hours).

A weather-informed flight and airspace optimization framework that jointly optimizes individual flight routes and multi-flight airspace utilization to minimize the combined warming effects of CO2 and non-CO2 emissions. Summary results using historical flight data will be presented to demonstrate the scaled impact of navigational avoidance strategies.

Together, these two research efforts represent a significant step toward reducing aviation's overall climate impact through advanced predictive modeling and operational optimization.

2 - Aircraft Engine Maintenance, Repair, and Overhaul Operations Optimization at GE Aerospace

Sohum Chatterjee, GE Aerospace, Niskayuna, NY, United States, Israel Beniaminy, David Toledano, Onur Dulgeroglu, Banu Gemici-Ozkan

Aircraft engine repair is a complex, multi-site operations management challenge that demands precise coordination across various stages and locations. After induction into the overhaul shops, each engine undergoes a detailed inspection and workscoping process to identify the necessary repairs or replacements. The core challenge lies in ensuring that all repair operations are completed efficiently and within target deadlines, enabling the timely return of engines to airline. This process requires periodic decision-making across multiple organizational levels, ranging from tactical day-to-day operations to strategic long-term planning, while balancing competing objectives. A complicating factor is the inherent uncertainty in the network, such as repair and supplier lead time variation or process yields that may require rework of certain parts.

In this work, we present GE Aerospace's approach to tackling these challenges, with a specific focus on optimizing the process of part allocation to replace scrapped components in overhaul engines. Part allocation is a critical process that ensures the optimal use of the spare part pool by allocating the right quantities to the right engines at the right time. We develop a multi-objective mixed-integer programming model and generate solutions that significantly improve fleet repair performance over current allocation practices. Preliminary results illustrate improved decision-making efficiency, reduced launch delays, and significantly enhanced throughput rates.

3 - Cogeneration Power Plant Dispatch Problem

Rajesh Tyagi, Retired, Plainsboro, NJ, United States

In a cogeneration power plant, an assembly of gas turbines, steam turbines and other auxiliary units such as boilers and duct burners work in tandem to produce electric power and steam. The operational settings that minimize fuel cost for meeting energy demand requirements, possibly over multiple cogeneration blocks, entails diligent adjustments across settings and blocks, whose combinatorics favor a complex optimization solution. We describe the problem and suggest a practical solution approach.

4 - Optimal Scheduling of Field Resources for Power Plant Outages

Rajeev Namboothiri, GE Vernova Research, Bangalore, India

GE Vernova's Gas Power Business installs and supports power generation equipment around the world. These power plants undergo periodic planned outages and forced unplanned outages to ensure safe and efficient operations. These outages require mobile field resources with specific skillset for a specific duration to perform maintenance tasks. A host of additional constraints restrict the eligibility and availability of a field resource to perform a maintenance task. The Business centrally manages the deployment of field resources to power plant outages through five different regions. We developed a Field Resource Optimizer product to provide resource scheduling recommendations to minimize the total cost associated with field service operations while meeting all the real-world constraints associated with assigning field resources to outage tasks. The standard optimization formulation of the field resource allocation problem results in a mixed integer linear program that is too large to be solved in a reasonable amount of time for practical use. We therefore developed a novel solution procedure that iteratively solves a series of linear programs to achieve the optimal solution within a few minutes. The Business is actively using this tool since 2024, resulting in significant operational cost savings.

5 - Wind Turbines Services Optimization

Srinivas Bollapragada, GE Vernova Advanced Research Center, Niskayuna, NY, United States

This work focuses on developing advanced optimization algorithms and software tools to enhance the servicing of onshore wind turbines. The primary goals are to improve customer satisfaction, reduce service costs, and strategically position the services organization for the anticipated significant growth in the installed base of onshore wind turbines. To achieve these objectives, we have designed and implemented a suite of optimization solutions and digital tools that are currently deployed at GE Vernova, delivering measurable improvements in service operations.

6 - Is Data-Driven Failure Early Warning Possible for Wind Turbines?

Yu Ding, Georgia Institute of Technology, Atlanta, GA, United States

Unexpected failures in engineering systems lead to expensive maintenance actions and should be avoided if at all possible. This is particularly true for wind turbine systems, for which unexpected failures not only demand costly repairs but also cause long downtime. Currently early warning systems work based on physically verified symptoms, much like in human illness early warning. A domain expert identified a symptom (like a crack), which, once initiated, will eventually lead to failure. When the enabling domain knowledge is available, an early warning problem can be converted to a regular change or anomaly detection problem. The limitation is that practitioners do not always know which symptoms to look for. It is also rather costly to thoroughly screen the whole turbine system on a frequent basis for the onset of certain damage inception conditions. On the other hand, modern wind turbines (in fact for most modern engineering systems) are equipped with a range of sensors, forming a system known as SCADA (supervisory control and data acquisition) that collects data continuously. The speaker wants to shed some lights on the question whether one can provide early-warning capability by just using the SCADA data.

MD18

Building A Level 4 A404

Health and Holistic Fitness: Using Big Data from Wearables, Free Text, and Clinical Sources to Make Personalized Decisions

Invited Session

Military and Security

Chair: Russell Nelson, United States Military Academy, West Point, NY, United States

Co-Chair: Diana Thomas, United States Military Academy, West Point, NY, United States

1 - Mining Pubmed with LLMs: Enhancing Literature Synthesis with AI

Jonathan Day, United States Military Academy, West Point, NY, United States

We conducted a large-scale natural language processing (NLP) analysis of 7,366 medical research articles related to precancerous breast lesions, including lobular proliferative hyperplasia (LPH), atypical ductal hyperplasia (ADH), lobular carcinoma in situ (LCIS), flat epithelial atypia (FEA), and other atypical proliferative lesions. Articles were retrieved using a hybrid web scraping framework combining BioEntrez and Selenium. To explore latent themes and synthesize findings across the literature, we generated document embeddings using both BERT-based contextual models and TF-IDF representations. These embeddings were then analyzed through unsupervised learning methods, including clustering algorithms (e.g., K-Means, DBSCAN) and pairwise similarity metrics (cosine similarity and alternative distance measures), to produce interpretable visualizations of thematic groupings. Additionally, we employed abstractive summarization models such as BART and PEGASUS to distill clinical insights, with a specific focus on statistical measures commonly reported in epidemiological studies—odds ratios (OR), hazard ratios (HR), and relative risks (RR). Our analysis offers a novel approach in achieving objective, reproducible synthesis of complex and heterogeneous literature. Our machine learning-driven framework demonstrates the potential for automating literature triage, identifying emergent subtopics, and enhancing interpretability through scalable, data-driven summarization. These findings suggest that NLP tools can augment the systematic review process, enabling a more nuanced and statistically enriched understanding of the evolving field of precancerous breast pathology.

2 - From Device to Model: Extracting Insight From Cgm Metrics

Austin Semmel, North Carolina State University, Raleigh, NC, United States

Despite their name, continuous glucose monitors (CGMs) face the reality of discretized monitoring intervals, which presents an inherent tradeoff between measurement frequency and device longevity due to battery constraints. This limitation creates challenges for extracting insights from CGM data. Missing data points and varied reading intervals both significantly impact the consistency of derived glycemic metrics, making it difficult to compare results across studies and inform personalized readiness strategies. We developed a Python-based application that addresses these issues by accepting raw CGM data from any device, performing automated preprocessing, and calculating over 15 validated metrics. The tool emphasizes transparency and reproducibility with standardized metric definitions and user-adjustable parameters for handling missing data. We demonstrate the importance of considering data completeness thresholds and reading intervals (5, 10, 15 minutes, etc.) when assessing metric consistency, as these factors substantially influence metric values and their subsequent interpretation. The tool's output files integrate with statistical modeling and machine learning applications to drive inferences from CGM data. CGM data does not lend itself to "one-size-fits-all" analysis. Harnessing this mindset, we provide a standardized, multi-metric approach to CGM data analysis. Our tool supports clinical decision-making and large-scale research initiatives aimed at understanding patterns of both glycemic control and variability. This work represents a critical step towards harnessing the potential of CGM technology to inform personalized strategies for optimizing individual glycemic control and overall health, ultimately contributing to improved readiness across the force.

3 - The Chemistry of Health: How Metabolomics is Redefining Precision Medicine

Blake Rushing, Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC, United States

Metabolomics, the comprehensive study of small molecules or metabolites in biological systems, offers a unique and powerful lens through which to study human health. As the “ome” closest to phenotype, the metabolome serves as a dynamic readout of gene activity, environment, diet, and lifestyle. These interactions shape metabolic pathways that regulate a wide array of biological outcomes such as energy production, immune response, detoxification of xenobiotics, gene expression, and much more. As such, metabolomics is increasingly recognized as an essential tool in biomedical research and precision health. This talk will introduce core concepts of metabolomics, including how high-resolution analytical techniques—primarily mass spectrometry—are used to capture thousands of features from biological samples like blood, urine, cells, and tissues. These data can be analyzed to uncover disrupted pathways, detect early signs of disease, design nutritional/pharmacological interventions, or evaluate responses to interventions such as diet, medication, or environmental exposures. Finally, we will consider the future of metabolomics in advancing precision nutrition and medicine. By integrating metabolic profiles with genetic and environmental data, researchers can better understand disease risk and response at the individual or subpopulation level. The ultimate goal is to translate this information into actionable strategies that promote health and prevent disease.

4 - Fusing Multi-Modal Health Data for Personalized Predictions: Toward Integrated Health Modeling

Jake LaPorte, United States Military Academy, West Point, NY, United States

As the number of digital health monitors increases and the cost of these devices decreases, it is possible for individuals to continuously monitor their health, activity level, and how their body interacts with food. A significant challenge comes with the sheer volume of physiological data associated with these devices in the form of turning this raw data into meaningful, actionable information for the user, along with practical insights into each person's health. The focus of this talk is the modeling technique used to transform accelerometer data into steps per day and peak-30 cadence using publicly available data, feature engineering, and a predictive, ensemble model that includes a classification model which feeds a regression model.

MD19

Building A Level 4 A405

High-Dimensional Stochastic Controls Problems in Operations Management

Invited Session

Applied Probability Society

Chair: Yuan Zhong, University of Chicago, Chicago, IL, United States

1 - Polynomial TIME Approximation Schemes for Online Resource Allocation with General Correlations

David Goldberg, Cornell University, Ithaca, NY, United States, Yilun Chen, Umur Cetin

There has been a growing interest in studying online resource allocation problems such as network revenue management under more general correlation structures, motivated by the complex data sets, generative models, and high-variance phenomena driving modern applications. Several interesting recent works either posit a particular structure for the possible correlations, have a complexity scaling with the number of Markovian “states of the world” (which may be exponentially large), or make additional continuity assumptions.

We show that the NRM problem (suitably normalized) has an additive Polynomial Time Approximation Scheme (with runtime independent of the number of “states of the world”) under a completely general model of correlations, assuming access to a natural Monte Carlo simulator for the demand dynamics, and that the maximum number of resources used by a product is uniformly bounded.

In contrast to several past works on PTAS for multi-stage stochastic linear programming with runtime exponential in the time horizon, our methods have a runtime essentially independent of the horizon altogether.

At the heart of our algorithms is a new way to conceptualize and implement stochastic gradient methods on-the-fly/recursively for the massive linear programs (on scenario trees, i.e. deterministic-equivalent problem) associated with natural relaxations of these problems, based on the simple idea that to implement a policy one needs only compute the relevant values along the sample path presented to you.

We also derive new algorithms and insights for several online combinatorial optimization problems with general correlation structure on bounded degree (random) graphs.

2 - Dynamic Scheduling of a Multiclass Queue in the Halfin-Whitt Regime: a Computational Approach for High-Dimensional Problems

Ebru Kasikaralar, University of Chicago, Booth School of Business, Chicago, IL, United States, Baris Ata

We consider a multi-class queueing model of a telephone call center, in which a system manager dynamically allocates available servers to customer calls. Calls can terminate through either service completion or customer abandonment, and the manager strives to minimize the expected total of holding costs plus abandonment costs over a finite horizon. Focusing on the Halfin-Whitt heavy traffic regime, we derive an approximating diffusion control problem, and building on earlier work by Beck et al. (2021), develop a simulation-based computational method for solution of such problems, one that relies heavily on deep neural network technology. Using this computational method, we propose a policy for the original (pre-limit) call center scheduling problem. Finally, the performance of this policy is assessed using test problems based on publicly available call center data. For the test problems considered so far, our policy does as well as or better than the best benchmark we could find. Moreover, our method is computationally feasible at least up to dimension 500, that is, for call centers with 500 or more distinct customer classes.

3 - A Computational Method for Solving the Stochastic Joint Replenishment Problem in High Dimensions

Wouter van Eekelen, The University of Chicago Booth School of Business, Chicago, IL, United States, Baris Ata, Yuan Zhong

We consider a class of stochastic joint replenishment problems in a high-dimensional setting. First, we approximate the problem by a multi-dimensional impulse control problem. Exploiting connections among the impulse control problem, backward stochastic differential equations (BSDEs) with jumps, and the stochastic target problem, we develop a simulation-based computational method that relies on deep neural network technology. We test our proposed policy against the best available benchmarks in the context of the original stochastic joint replenishment problem. For the test problems studied thus far, our method is as good or better than the best benchmark we could find, and it is computationally feasible up to at least 100 dimensions—that is, 100 stock-keeping units (SKUs).

4 - Multi-Agent Markov Entanglement

Shuze Chen, Columbia University, New York City, NY, United States, Tianyi Peng

Value decomposition has long been a fundamental technique in multi-agent reinforcement learning and dynamic programming. Specifically, the value function of a global state $(s_1, s_2, \dots, s_N)$ is often approximated as the sum of local functions: $V(s_1, s_2, \dots, s_N) \approx \sum_i V_i(s_i)$. This approach has found various applications in modern RL systems. However, the theoretical justification for why this decomposition works so effectively remains underexplored. In this paper, we uncover the underlying mathematical structure that enables value decomposition. We demonstrate that a Markov decision process (MDP) permits value decomposition *if and only if* its transition matrix is not “entangled”—a concept analogous to quantum entanglement in quantum physics. Drawing inspiration from how physicists measure quantum entanglement, we introduce how to measure the “Markov entanglement” and show that this measure can be used to bound the decomposition error in general multi-agent MDPs. Using the concept of Markov entanglement, we proved that a widely-used class of policies, the index policy, is weakly-entangled and enjoys a sublinear $O(\sqrt{N})$ scale of decomposition error for N -agent systems. Finally, we show Markov entanglement can be efficiently estimated, guiding practitioners on the feasibility of value decomposition.

5 - Multilevel Picard Approximations for Drift Rate Controls with State Constraints

Yuan Zhong, University of Chicago, Chicago, IL, United States

Motivated by applications to queueing network controls, we develop a multi-level Picard (MLP) approximation scheme for solving high-dimensional drift rate control problems whose states are constrained to stay within the nonnegative orthant. We prove that under suitable conditions, the MLP scheme provably overcomes the curse of

dimensionality. In our numerical experiments, the scheme is shown to be computationally feasible for problems up to dimension 20.

MD20

Building A Level 4 A406

Emerging Technologies for Railway Transportation

Invited Session

Railway Applications

Chair: Xingang (Clark) Cheng, RailTek, LLC, 5550 Tenbury Way, Johns Creek, GA, 30022, United States

1 - Assessing the Capacity of Railway Network Control Points with Monte Carlo Simulation

Jiayi Zhao, DB E.C.O. North America, San Gabriel, CA, United States, Darkhan Mussanov, Yasmin Ara

Evaluating the capacity of control points, especially in dense railway traffic regions such as St. Louis, is critical for strategic planning and congestion management. However, existing modeling tools often require detailed knowledge of both the physical layout of infrastructure elements (e.g., diamonds, leads from yards to mainlines, and junctions) and train schedules. This study introduces a simulation-based parametric methodology to estimate the capacity of these key control points without relying on detailed layout measurements or operational timetables. The approach begins by generalizing control point configurations into abstract movement patterns that capture common train interactions such as crossings, merges, and divergences. Monte Carlo simulation is then used to model train movements through these control points over defined time intervals, capturing stochastic occupancy patterns and conflict dynamics. Capacity is quantified in terms of the number of feasible train movements per time period under varying traffic conditions. Moreover, multiple control points can be aggregated to represent larger, more complex network regions, enabling evaluation of cumulative capacity. This flexible, data-light approach supports scalable assessments across diverse network topologies and operational uncertainties. The results provide critical insights into throughput limitations in high-density corridors and can inform decisions on bottleneck mitigation, infrastructure investment, and strategic rail planning. By abstracting physical and temporal specifics, the method enables early-stage capacity analysis and comparative evaluations across urban rail systems. This work contributes a practical tool for rail network operators, planners, and researchers seeking efficient, layout-independent methods for capacity evaluation.

2 - An Outer-Inner Approximation Method for the Generic Choice-Based Optimization Problem

Haoye Chen, KTH Royal Institute of Technology, Stockholm, Sweden, Jan Kronqvist, Ma Zhenliang

Choice-based optimization problem integrates demand modeling into optimal supply decisions, which is generic for decision-making applications. Solving the problem is challenging given the nonlinear discrete choice model constraints. Existing solution methods are limited to specific problem structures, such as binary or discrete supply decisions and fixed option attributes. This paper proposes an outer-inner approximation method for the generic choice-based optimization problem without specific problem structural requirements. We validated the method using a network expansion problem on the Sioux Falls network, aiming to reduce the overall system congestion by optimally expanding road capacities considering the road expansion cost. The results show that the expansion cost is significantly lower than the total travel time savings. More experiments are expected to benchmark with existing models using more case studies, e.g., service frequency and pricing in multimodal transportation systems.

3 - OneBOM: Accelerating Railcar Quoting and Improving Supply Chain Accuracy with AI and Analytics

Garrett Mucha, Trinity Industries, Dallas, TX, United States

Trinity Industries has developed OneBOM, an AI-powered platform that transforms how the company estimates costs, builds quotes, and aligns supply planning for complex, custom-engineered freight railcars. Built in Palantir Foundry, OneBOM consolidates five engineering BOM sources into a unified data model that supports rapid, accurate quoting and provides a consistent foundation for supply chain planning.

A key innovation is the integration of generative AI to bridge engineering specifications and BOM creation. Trinity's "Spec-to-Spec diff" capability compares incoming customer specs against historical builds, highlighting key design changes and mapping them directly to affected BOM elements. This allows estimators to bypass manual BOM construction and instead generate quote-ready BOMs directly from specification changes—reducing cycle time and improving consistency.

OneBOM also features advanced cost modeling that calculates material, labor, and overhead in real time, allowing estimators to understand cost impacts instantly. Alternate parts logic and reusable templates further speed up the quoting process while preserving flexibility and accuracy.

On the supply side, BOM harmonization improves forecast precision. Cleaned and structured BOMs now feed reliable demand signals into procurement models, improving long-lead material planning and reducing inventory risk.

With OneBOM, Trinity has cut quoting effort by 80%, improved supply chain forecast accuracy, and enabled a more agile response to customer-specific requirements. The platform showcases how GenAI and analytics can modernize railcar manufacturing by aligning engineering data, cost estimation, and procurement into a unified digital workflow.

4 - Comparative Analysis of Machine Learning and Traditional Methods in Predicting Railroad Operational Metrics

Muhammad Ramahi, TTX Company, Charlotte, NC, United States

We present an in-depth exploration of a novel technique for forecasting critical railroad operational metrics, specifically focusing on cycle-time and railcars network mileage. The primary objective is to compare and contrast the efficacy of emerging machine learning techniques with traditional classical methods in predicting these metrics for various fleets.

Cycle time, which refers to the total time taken for a railcar to complete a full cycle from departure to return, is a crucial component of the business planning process.

Accurate prediction of cycle-time is essential as it helps determine the number of railcars required to meet network demand efficiently. An accurate forecast ensures that there are neither too many idle railcars nor too few to meet the demand, thereby optimizing operational efficiency and cost-effectiveness.

On the other hand, predicting railcars network mileage, which is the total distance covered by railcars within the network, is vital for planning and budgeting equipment maintenance. Accurate mileage predictions enable better scheduling of maintenance activities, ensuring that railcars are serviced at appropriate intervals to prevent unexpected breakdowns and extend their operational lifespan. This not only enhances the reliability and safety of the rail-network but also helps in optimizing maintenance budgets.

By leveraging machine learning techniques, which can analyze vast amounts of data and identify complex patterns, we aim to improve the accuracy and reliability of these forecasts. This study will provide valuable insights into the potential advantages and limitations of machine learning compared to classical methods, ultimately contributing to more efficient and effective railroad operations.

5 - Dynamic Routing for Intermodal Container Transportation

Qianqian Tong, University of Texas at Austin, Austin, TX, United States, Tyler Dick

North America's rail-road intermodal container transportation system handles over 18 million units annually. With the rapid growth of container shipment demand across the country, efficient routing methods within and between intermodal terminals play a critical role in achieving both strategic and operational goals. This study assigns daily container flows across intermodal routes by minimizing weighted travel time, cost, and energy per container to support planning-level mode choice decisions for container transportation. To consider congestion and delay effects in routing and capacity planning, this study also proposes a two-layer simulation-based framework for container route choice. At the node level, we derive node performance functions (NPFs) using SimPy-based discrete-event simulation, where optimal layout and vehicle performance parameters are used to evaluate how terminal yard scale and resource combinations influence container processing efficiency. At the link level, we apply the Advanced Locomotive Technology and Rail Infrastructure Optimization System (ALTRIOS) to simulate and obtain link performance functions (LPFs) to capture the dynamic performance of mainlines, and also apply the Bureau of Public Roads (BPR) function to describe the performance of adjacent roadways under traffic. Integrating NPFs and LPFs within a dynamic traffic assignment framework, we merge them to create a dynamic container routing model to provide strategic guidance for intermodal

network design and operations planning. To demonstrate the applicability of our approach, we implement our model in a Texas-based case study to examine how strategic and tactical decisions to improve efficiency can be made regarding both vehicle technologies and infrastructure configurations.

MD21

Building A Level 4 A407

TSL Best Student Paper Award

Award Session

Transportation Science and Logistics (TSL)

Chair: Okan Arslan, HEC Montreal, Montréal, QC, Canada

Co-Chair: Baris Yildiz, Koc University, Rumelifeneri Yolu, Istanbul, 34450, Turkey

1 - Overworking or Outsourcing: Integrated Prediction- Optimization for Human-Centric Last-Mile Delivery Zone Assignment

Haoran Guo, Tsinghua University, Beijing, China, People's Republic of

2 - Column Elimination: An Iterative Approach to Solving Arc Flow Formulations

Anthony Karahalios, OptiLogic, Warrendale, PA, United States

Coming soon

3 - Conformal Inverse Optimization for Adherence Aware Prescriptive Analytics

Bo Lin, University of Toronto, Toronto, ON, Canada

Inverse optimization has been increasingly used to estimate unknown parameters in an optimization model based on decision data. We show that such a point estimation is insufficient in a prescriptive setting where the estimated parameters are used to prescribe new decisions. The prescribed decisions may be low-quality and misaligned with human intuition and thus are unlikely to be adopted. To tackle this challenge, we propose conformal inverse optimization, which seeks to learn an uncertainty set for the unknown parameters and then solve a robust optimization model to prescribe new decisions. Under mild assumptions, we show that the suggested decisions can achieve bounded out-of-sample optimality gaps, as evaluated using both the ground-truth parameters and the decision maker's perception of the unknown parameters. Our method demonstrates strong empirical performance compared to classic inverse optimization.

4 - A Double Decomposition Algorithm for Network Planning and Operations in Deviated Fixed-route Microtransit

Alexandria Schmid, MIT, Cambridge, MA, United States

MD22

Building A Level 4 A408

Advances in Stochastic Network

Invited Session

Applied Probability Society

Chair: Lun Yu, Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A

1 - Redundancy OR Flexibility? Interplay Between Their Values in Designing Fork-Join Systems

Chutong Gao, Northwestern University, Evanston, IL, United States, Seyed Iravani, Ohad Perry

We consider fork-join systems with redundancy and flexibility. Fork-join systems are widely employed in online service platforms, among other real-world applications. In a k -requirement fork-join (FJ_k , $k \geq 2$) system, a job "forks" (i.e., divides) into k tasks which are processed in parallel stations; after completion, these k tasks then "join" together to finish the job. To minimize job sojourn times in FJ_k systems, two strategies---redundancy and flexibility---are often employed. Redundancy deploys $n \geq k$ stations and sends one task to each server, but only k out of n tasks are needed to complete each job. Flexibility enables dynamic service capacity allocation across stations at any time. To understand the value of each strategy and their interplay in reducing sojourn times, we propose a unified modeling framework that simultaneously considers: (i) how many redundant stations should be established, and (ii) how the service capacity should be allocated among all stations. With flexibility, we characterize the optimal dynamic capacity-allocation policy, under which redundancy does not offer any additional value in FJ_k systems. In contrast, in the absence of flexibility, for $k=2$, we prove "a little redundancy goes a long way" in a twofold sense: To achieve asymptotic optimality relative to the best system with flexibility, it suffices to (i) employ only one extra station, making $n=k+1$, and (ii) allocate an arbitrarily small capacity $\epsilon > 0$ to that extra station. This suggests that (even a little) redundancy is as attractive as (full) flexibility. Extensive numerical studies further extend the insights to general FJ_k systems.

2 - A near-Optimal Control Policy for Data-Driven Assemble-to-Order Systems

Zhixuan Cai, Tsinghua University, Beijing, China, People's Republic of, Lun Yu, Zhaoran Wang, Tianhu Deng

We study a data-driven control of assemble-to-order (ATO) systems, aiming to synchronize between components' ordering decisions under non-identical lead times and products' assembly under unknown demand distribution. To the best of our knowledge, our work is the first to analyze the sampling effectiveness of general ATO system as high-dimensional control problem on noncompact domain and the first to propose an algorithm that finds a near-optimal solution. Surprisingly, we prove that the data-driven ATO control problem does not suffer from the curse of dimensionality, in the sense that the performance gap between the best data-driven control policy and the intractable optimal policy is only logarithmically larger than the gap between sample mean and expectation. With a novel utilization of an important convex property of ATO transition dynamics, we develop a specialized reinforcement learning (RL) algorithm that is applicable to the control problem, which also provably achieves the best performance gap. An extensive numerical study supports that a close-to-optimal policy of the general ATO system is obtainable using small demand samples, where the proposed algorithm consistently and notably outperforms existing heuristic policies and an off-the-shelf RL algorithm.

3 - Functional Limits of Generalized Jackson Networks in Multi-Scale Heavy Traffic

Jin Guang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Zhen-Qing Chen, Jim Dai

We investigate the process-level convergence of generalized Jackson networks (GJNs) in multi-scale heavy traffic recently proposed by Dai, Glynn, and Xu (2023), which yields a limit with a product-form stationary distribution. Although this regime effectively approximates the stationary distribution, it does not provide insights into the temporal evolution of the process or explain why each component becomes independent. In this paper, we demonstrate that a sequence of GJNs converges weakly to a limiting process with independent components, offering a deeper understanding of their temporal dynamics and independence structure. Additionally, we extend the multi-scale heavy traffic regime to a block-scale heavy traffic regime, where the components are divided into blocks, and each block is scaled at the same rate. We prove that process-level convergence holds in this regime as well, and the limiting process exhibits block-wise independence. Our results provide a new perspective on the temporal behavior of GJNs in multi-scale and block-scale heavy traffic regimes.

4 - Electrifying the Fleet: Dimensioning Electric Vehicle Sharing Systems

Jingyuan Wan, University of Michigan, Ann Arbor, MI, United States, Saif Benjaafar, Martin Zubeldia

We consider a system of shared electric vehicles that operates on an on-demand basis over a spatial network of multiple locations. Customers arrive at each location and, if there is a vehicle available, they travel to another location where the vehicle is dropped off. At each location, there is a finite number of chargers that can be used by the vehicles. We model this system as a closed queueing network and derive closed-form lower and upper bounds on the minimal fleet size and number of chargers at each location so that the fraction of demand unfulfilled and average waiting time for charging are below predetermined target service levels. The upper bounds imply

dimensioning rules that guarantee the target service levels. We show that the dimensioning rules implied by these bounds are asymptotically optimal. More significantly, these dimensioning rules are highly interpretable with separable components that can be attributed to different drivers of capacity. We leverage the upper and lower bounds to propose an approximation of the optimal capacity that is always between the upper and lower bounds and is asymptotically optimal and otherwise exhibits strong performance as validated by extensive numerical experiments. We show how the results we obtain can be used to examine the tradeoff between fleet size and size of charging infrastructure and assess the impact of electrification.

5 - Cascading Bandits for Periodic Recommendations: Handling Delayed Feedback and User Retention

Dairui Wang, Tsinghua University, Beijing, China, People's Republic of, Junyu Cao, Wei Qi

Delayed feedback is a critical problem in dynamic recommender systems. In practice, the feedback result often depends on the frequency of recommendation. Most existing online learning literature fails to consider optimization of the recommendation frequency, and regards the reward from each successfully recommendation as equal. In this work, we consider a novel cascading bandits setting, where individual messages from a selected list are sent to a user periodically. Whenever a user does not like a message, she may abandon the system with a probability positively correlated with the recommendation frequency. A learning agent needs to learn both the underlying message attraction probabilities and users' abandonment probabilities through the randomly delayed feedback. We first show a dynamic programming solution to finding the optimal message sequence in deterministic scenarios, in which the reward is allowed to vary with different messages. Then we propose a polynomial time UCB-based offline learning algorithm, and discuss its performance by characterizing its regret bound. For the online setting, we propose a learning algorithm which allows adaptive content for a given user, and its regret bound is also discussed. Numerical experiment on NetEase Cloud Music dataset confirms the effectiveness of our algorithms, and clarifies the value of contextual information in recommender systems.

MD23

Building A Level 4 A409

MIF Early Career Award and MIF Fellows

Award Session

Minority Issues Forum

Chair: Wesley Marrero, Thayer School of Engineering at Dartmouth, Hanover, United States

Co-Chair: Gian Garcia

1 - Towards Inclusive and Human-Centered AI: Research and Service at the Intersection of Algorithms and Society

Juba Ziani, Georgia Tech, Atlanta, GA, United States

Research-wise, my work reframes fairness not as a purely technical constraint but as a property that emerges from the broader socio-economic processes in which AI and decision making is deployed. I will briefly discuss how my research draws on tools from computer science, operations research, and economics to study short-term and long-term disparities and the social impact of algorithmic decisions. Service-wise, I have been striving to build structures that support the development of young researchers and inclusivity in our research communities. I will highlight some of these efforts—including founding the ISyE Rising Stars Workshop, leading the EAAMO Doctoral Consortium for the past 4 years, and developing mentoring and outreach programs aimed at broadening participation not only in OR but also in computing.

2 - Bringing the Energy Transition to Underserved Texas Communities: the City of Dallas Green Warriors Weatherization Pilot

Erick Jones, University of Texas at Arlington, Arlington, TX, United States

Various federal, state, and local programs aim to improve energy outcomes for underserved low-income households. They mainly target the inefficient housing envelopes prevalent but new efforts have aimed to incorporate energy generation and storage as well. Nonetheless, administrative not technical hurdles provide the most resistance to successful deployment of these programs. Organizations design programs to address a specific challenge and often find other factors that are not budgeted for affect the efficacy of their targeted intervention. This either results in interventions that are less effective than originally planned or a shrinking pool of potential candidates. These challenges frustrate residents as the promised benefits never arrive even filling out an exhaustive level of paperwork and waiting years. To address this issue our team created a pilot program in South Dallas, a majority Black and low-income neighborhood, to weatherize homes using local contractors (Green Warriors) that overcomes previous barriers to success. To illustrate the efficacy of the program we collected data before and after the weatherization and are creating a replicable framework to be used for future projects.

3 - TBD

4 - TBD

MD24

Building A Level 4 A410

Two New Applications of the Gittins Index: Tail Scheduling and Bayesian Optimization

TutORial session

TutORials

Chair: Nilay Noyan, AMAZON, Bellevue, WA, United States

1 - The Gittins Index: A Design Principle for Decision-Making Under Uncertainty

Ziv Scully, Cornell University, Ithaca, NY, United States, Alexander Terenin

The Gittins index is a tool that optimally solves a variety of decision-making problems involving uncertainty, including multi-armed bandit problems, minimizing mean latency in queues, and search problems like the Pandora's box model. However, despite the above examples and later extensions thereof, the space of problems that the Gittins index can solve perfectly optimally is limited, and its definition is rather subtle compared to those of other multi-armed bandit algorithms. As a result, the Gittins index is often regarded as being primarily a concept of theoretical importance, rather than a practical tool for solving decision-making problems.

The aim of this tutorial is to demonstrate that the Gittins index can be fruitfully applied to practical problems. We start by giving an example-driven introduction to the Gittins index, then walk through several examples of problems it solves—some optimally, some suboptimally but still with excellent performance. Two practical highlights in the latter category are applying the Gittins index to Bayesian optimization, and applying the Gittins index to minimizing tail latency in queues.

MD25

Building A Level 4 A411

Coordinated Collaboration and Logistics Risk Management

Invited Session

TSL: Freight Transportation

Chair: Ilke Bakir, University of Groningen, Groningen, Netherlands

1 - A New Perspective on Carrier Collaboration: Collaborative Vehicle Utilization

Ilke Bakir, University of Groningen, Groningen, Netherlands, Sahand Asgharieh Ahari, Kees Jan Roodbergen

We study a centralized collaboration mechanism for logistics carriers based on the shared use of vehicle fleets, which we refer to as Collaborative Vehicle Utilization (CVU). While most collaboration frameworks focus on order sharing, where carriers deliver on each other's behalf, CVU enables carriers to lend and borrow vehicles

while retaining full control over their own delivery operations. This avoids the need to share sensitive customer or operational data, making collaboration more viable in practice. CVU also supports sustainability efforts by improving vehicle utilization and reducing the need for fleet expansion, which is particularly relevant given the environmental footprint of vehicle production and the limited scalability of electric truck manufacturing in the coming years. We explore CVU in three collaborative vehicle routing settings. For each setting, we develop formulations, branch(-cut)-and-price methods, and heuristics. In the single-period setting, we compare CVU to order sharing and non-collaborative operations and find that CVU achieves similar profits while offering substantial computational advantages. In a multi-period extension, where customers require consistent delivery time windows, CVU reduces network-level fleet size by up to 45% and increases profit by up to 15%. Finally, we apply CVU to a two-echelon urban logistics system, where large trucks operate from distribution centers to satellites and smaller vehicles deliver to customers, potentially through multiple trips starting and ending at the satellites. In this variant, carriers may share vehicles in both echelons and optionally share satellite capacity, which further improves vehicle utilization and yields additional profit gains.

2 - Improving Supply Chain Resilience with Information Sharing Through a Central Platform

BATUHAN Celik, Rensselaer Polytechnic Institute, Troy, NY, United States, Bahar Cavdar, Jennifer Pazour

The ability of supply chain players to adapt and change decisions in response to disruptions is key to increasing resiliency. However, disruptions are observed asynchronously, reducing the supply chain's collective response and amplifying disruption impacts. These effects can be alleviated through information sharing, but firms hesitate due to strategic risks. To address this, we design an information-sharing policy for competing firms to improve supply chain resilience. A key characteristic of the designed policies is to ensure that players maintain decision autonomy and competitive power. We explore how sharing disruption information affects profit-oriented equilibrium strategies and resilience, using a central platform. We aim to design a platform that determines optimal information-sharing policies to maximize resilience while players act according to profit-oriented objectives. We use a game-theoretic model of a two-echelon supply chain with distributors and retailers. We analyze when resilience aligns or conflicts with personal objectives and develop a mechanism to incentivize players to participate in the platform, namely committed sharing. We examine the impact of network structure, disruption type, player characteristics and participation on resilience and profit under this sharing approach. We identify which distributors' participation benefits resilience and set platform parameters for large-scale supply chains with an optimization model. We show how supply chain characteristics impact sharing behavior and resilience, offering insights for policymakers and academics designing platforms that consider system-wide objectives while improving decision-making under uncertainty.

3 - A Mean-Variance Optimization Model for Produce Export Distribution with Postponed Shipment and Market Allocation Option

Alejandro Mac Cawley, Pontificia Universidad Catolica de Chile, Santiago, Chile, Nicolás Vargas, Mathias Klapp

This research addresses the challenge of managing price risk in the allocation decisions of exporters of perishable agricultural products, such as grapes or apples, sold in geographically dispersed spot markets characterized by volatile and correlated prices and long transit times. We propose a methodology that extends the classic Markowitz Mean-Variance (MV) model, presenting a two-stage stochastic optimization framework that maximizes the operation's net income subject to a maximum acceptable income variability. Traditional risk management and optimization models often overlook diversification and recourse risk management strategies. The proposed model allows decision-makers to use three risk-mitigation strategies: distribute shipments across multiple markets (inter-market diversification), distribute shipments in multiple moments, and delay a fraction of their shipment allocation decisions to a later stage (inter-temporal diversification). Diversifying shipments across multiple markets and shipment moments allows the exporter to exploit negative price correlations, reducing volatility for a given expected income. Delaying shipment allocation decisions allows the decision-maker to rely on updated price information, reduce uncertainty, and increase expected net income for a given level of risk tolerance. The study demonstrates the value added by combining inter-market, multiple shipment moments, and inter-temporal diversification strategies, especially when product inventory costs are manageable.

MD26

Building A Level 4 A412

Applications of Agent-Based Modeling and Simulation in Decision Analytics for Public Services

Invited Session

Simulation

Chair: Yuan Zhou, University of Texas at Arlington, Arlington, TX, United States

1 - Modeling Criminal Behavior in Urban Environments Using Criminological Theory and Agent-Based Simulation

Shohreh Moradi, The University of Texas at Arlington, Arlington, TX, United States, Yuan Zhou, Victoria Chen, Jay Rosenberger, Christi L Gullion, Chen Kan

Understanding the dynamics of urban crime requires theoretical frameworks that capture the interplay between individual behavior, routine activity, and environmental context. A theory-driven approach to modeling criminal behavior via an agent-based simulation is presented, grounded in three key criminological theories: Routine Activity Theory (RAT), Rational Choice Theory (RCT), and Crime Pattern Theory (CPT). These theories inform the decision-making rules and interactions among offenders, law enforcement officers, and citizens within the model. RAT provides the foundational structure for identifying potential crime events, emphasizing the convergence of motivated offenders, suitable targets, and the absence of capable guardians. RCT shapes offender behavior by modeling decisions based on perceived costs and benefits. CPT introduces spatial logic by anchoring agent routines and interactions within urban activity nodes and pathways. Together, these theories enable the generation of realistic, emergent crime patterns from micro-level agent decisions. By embedding these frameworks in the simulation architecture, the model serves both as a testbed for theoretical exploration and a foundation for practical applications in crime prevention. The results demonstrate the value of theory-informed agent rules in reproducing observed spatiotemporal crime distributions and offer new insights into the behavioral underpinnings of urban crime.

2 - Evaluating Police Patrol Strategies Through Agent-Based Modeling of Urban Crime

Yuan Zhou, University of Texas at Arlington, Arlington, TX, United States

Effective policing strategies require methods/tools that can simulate the dynamic relationship between law enforcement actions and criminal behavior in urban environments. An agent-based model (ABM) was developed to evaluate the impact of different police patrol strategies on the incidence and distribution of urban street crimes, including robbery, burglary, and larceny. The model integrates real-world data, such as crime statistics, environmental layouts, and census demographics, to construct a realistic simulation of urban mobility and routine activity. Two patrol strategies are implemented and tested: hotspot policing where officers are deployed to areas with historically high crime rates, and randomized patrols in which officers with general responsibilities dedicate a percentage of their time to high-crime areas. The ABM captures how offenders adapt their behavior in response to police presence, revealing how patrol tactics influence crime deterrence. Simulation results indicate that hotspot patrols can significantly reduce crime. This finding highlights the strengths and trade-offs of each strategy and underscore the potential of ABM as a decision support tool for law enforcement agencies seeking optimal patrol allocation strategies and crime prevention outcomes.

3 - Experimental Study of Vehicle Routing for a Police Patrolling Agent Based Simulation Model

Victoria Chen, The University of Texas-Arlington, Arlington, TX, United States, Shohreh Moradi, Darya Abbasi, Matine Ghasemloo, Jay Rosenberger, Yuan Zhou

Hot spots are patrol areas that have a high crime rate. Hot spot police patrolling has been studied in the criminal justice literature as an effective approach for reducing crime. Hot spot policing can be modeled via a vehicle routing optimization (VRO). The optimal route depends on the location and crime rate at hot spots. For an Arlington, Texas case study, robbery crime rates are simulated using an agent based model (ABM) based on crime theories to evolve criminal response to police patrol. The ABM simulation is computationally intensive. Rather than executing the ABM simulation many times to study the impact on VRO, a design and analysis of computer experiments approach is introduced to enable efficient study of the impact of different crime patterns on VRO solutions.

4 - An Integrated Mixed-Integer Programming Framework for Route Optimization and Maximum Coverage in Time-Dependent Patrol Scheduling

Jay Rosenberger, The University of Texas at Arlington, Arlington, TX, United States, Shohreh Moradi, Yuan Zhou, Victoria Chen

In this presentation, we tackle the challenge of optimizing police patrol routing by integrating real-time, time-dependent crime data into a dynamic, multi-visit scheduling framework that fully captures the spatial and temporal complexities of urban crime. Our approach is founded on a novel mixed-integer linear programming (MILP)

formulation that extends the classical vehicle routing problem by incorporating multiple visit indices per hotspot. This extension not only facilitates the determination of optimal spatial routes but also dynamically allocates patrol durations and schedules revisits based on evolving crime intensities. By directly linking activation decisions with routing constraints, our model ensures sequential, logical ordering of visits while adapting service times and revisit frequencies to meet operational demands. We evaluate the model performance using various instances and real-time datasets, demonstrating its capacity to continuously adjust patrol strategies in response to fluctuating hotspot conditions. The insights derived from our study yield actionable, adaptive multi-visit patrol schedules that enhance both the efficiency and responsiveness of policing in large, dynamic urban environments.

MD27

Building B Level 2 B201

Resilient Networks

Invited Session

Computing Society

Chair: Himadri Sen Gupta, Colorado State University Pueblo, 202 W Boyd St, Pueblo, CO, 73019, United States

Co-Chair: Andrés González, University of Oklahoma, Norman, OK, United States

1 - Integrated Crop and Cargo War Risk Insurance: Application to Ukraine

Ismael Mohammad Pour, University of Massachusetts Amherst, Amherst, MA, United States, Anna Nagurney, Borys Kormych

Conflicts and wars can significantly disrupt global agricultural supply chains, with severe impacts on food security worldwide. War risk insurance, on the other hand, can mitigate farmers' losses. In this paper, we use the theory of variational inequalities to construct a new multicommodity international trade network equilibrium model with exchange rates on general transportation networks joining supply markets with demand markets under different wartime scenarios. The model incorporates capacities on production and transportation, as well as potential commodity losses in transportation. We then propose formulae for integrated crop and cargo war risk insurance premiums that are supply market and commodity specific and incorporate them into the model, along with government subsidies for the premiums. A series of numerical examples, both illustrative and algorithmically solved, focusing on the ongoing war on Ukraine, reveal the critical role of integrated war risk insurance and government support in sustaining agricultural commodity trade flows and supply market prices as well as protecting the revenue of farmers.

2 - Perishable Goods on a Roll: Optimizing Logistic Distribution Under Uncertainty Based on Remaining Shelf Life

Mauricio Gamez, Massachusetts Institute of Technology, Cambridge, MA, United States, Christopher Mejia, Thiago Serra

An important task in today's food industry is the control of product quality throughout the food supply chain. One of the main reasons that accelerates the spoilage of fresh food is disruptions in the supply chain network, such as poor food handling in facilities, transportation, packaging, and others. Models are needed to determine the optimal supply chain configuration to maintain food freshness and thus reduce food waste/loss. In this paper, freshness is integrated into the supply chain network by tracking the remaining shelf life. We formulate a mixed-integer linear programming (MILP) model under uncertainty that accounts for the disruptions along the logistics network. To model and solve this problem, we use two-stage stochastic programming and solve it with a sample average approximation (SAA) method. To accelerate the convergence of the algorithm, we have developed a rolling horizon heuristic that speeds up the performance of the model to obtain near-optimal solutions. The resulting model is applied in an illustrative case study and can be used for the planning and operation of food distribution systems considering freshness and cost. The results show that the proposed approach can achieve high quality solutions with guaranteed performance.

3 - Fortifying Distribution Network Nodes Subject to Different Disruption Levels

Bahar Çavdar, Rensselaer Polytechnic Institute, Troy, NY, United States, Pelin Kesrit, Joseph Geunes

We consider a distribution network for delivering a natural resource or physical good to a set of nodes, each of which serves a set of customers, in which flow disruptions may occur at one or more nodes. Each node receives flow via a path from one or more source nodes, implying that a node's service level depends on flow from upstream nodes. We consider the effects of disturbances of an uncertain degree of severity that can lead to reduced flow at individual nodes in the network, as well as fortification measures for protecting nodes against these disturbances. Our proposed model determines node fortification levels that maximize service availability under a limited budget and various assumptions on network structure, the nature of potential disturbances, and their associated probability distributions, which may be different at each node.

4 - Leveraging Decoupling in Robust Network Rebalancing

Andre Cire, University of Toronto, Toronto, ON, Canada, Aliaa Alnaggar, Adam Diamant

Many critical applications in resource sharing, humanitarian logistics, and healthcare can be modeled as two-stage distributionally robust optimization problems, where the second stage involves rebalancing supply across a network. For example, in healthcare systems, medical equipment or staff often must be reallocated in response to uncertain demand across hospitals. Existing methodologies for such problems are largely based on linear decision rules and heuristics, which can be suboptimal and computationally demanding for large-scale or highly uncertain environments.

In this work, we propose an exact approach for rebalancing problems with discrete, separable ambiguity sets, such as two-moment and polyhedral uncertainty sets. Specifically, we exploit an inherent decoupling property of the optimal second-stage network flow solutions to design a novel single-stage reformulation based on network models and decision diagrams. The resulting formulation is exponentially smaller than classical robust optimization models and enables new combinatorial separation techniques with stronger convergence.

We evaluate the approach using a real-world case of resource distribution in a hospital network, presenting optimal distributionally robust solutions for large-scale instances. In particular, we assess the value of optimality in identifying resilient yet less costly robust solutions, also comparing with traditional approaches such as linear decision rules.

5 - Multi-Objective Multi-Agent Coordination for Hurricane Preparedness and Response

Jessica Heier Stamm, Kansas State University, Manhattan, KS, United States, Ayesha Farooq

A multi-objective multi-agent coordination framework is presented, and its application is demonstrated using a case study approach for facility location and distribution in hurricane response. The framework integrates cooperative game theory and goal programming, enabling the evaluation of both operational and strategic aspects of humanitarian logistics. The analysis highlights that, with minor adjustments, fair distribution of costs and demands is attainable, emphasizing the stability and sustainability of inter-agency coordination. The framework ensures an equitable demand fulfillment across all affected areas while maintaining near-optimal cost efficiency for the responding organizations. Moreover, the analysis reveals that coordinated strategies, achieved through resource pooling among autonomous organizations, lead to a more effective network of facilities compared to independent actions. Overall, the findings suggest that integrated resource management not only enhances operational responsiveness and service levels in hurricane scenarios but also delivers superior returns on investment, offering valuable insights for policymakers and disaster response planners.

MD28

Building B Level 2 B202

Experimentation and Decision-making

Invited Session

Computing Society

Chair: Amine Bennouna, Kellogg, Northwestern University, Cambridge, MA, United States

Co-Chair: Lin Fan, Northwestern University, Kellogg School of Management, Evanston, IL, United States

1 - How Much Data Do You Need? An Exact Characterization for Linear Optimization**Omar Bennouna, MIT, Cambridge, MA, United States**, Amine Bennouna, Saurabh Amin, Asuman Ozdaglar

We study the fundamental question of how informative a dataset is for solving a given decision-making task. In our setting, the dataset provides partial information about unknown parameters that influence task outcomes. Focusing on linear programs, we characterize when a dataset is sufficient to recover an optimal decision, given an uncertainty set on the cost vector. Our main contribution is a sharp geometric characterization that identifies the directions of the cost vector that matter for optimality, relative to the task constraints and uncertainty set. We further develop a practical algorithm that, for a given task, constructs a minimal or least-costly sufficient dataset. Our results reveal that small, well-chosen datasets can often fully determine optimal decisions—offering a principled foundation for task-aware data selection.

2 - Fast Revenue Maximization**Omar Besbes, Columbia University, New York, NY, United States**, Achraf Bahamou, Omar Mouchtaki

We study a data-driven pricing problem in which a seller offers a price for a single item based on demand observed at a small number of historical prices. Our goal is to derive precise evaluation procedures of the value of the historical information gathered by the seller, along with prescriptions for more efficient price experimentation. Our main methodological result is an exact characterization of the maximin ratio (defined as the worst-case revenue garnered by a seller who only relies on past data divided by the optimal revenue achievable with full knowledge of the distribution of values). This result allows to measure the value of any historical data consisting of prices and corresponding conversion rates. We leverage this central reduction to provide new insights about price experimentation. Motivated by practical constraints that impede the seller from changing prices abruptly, we first illustrate our framework by evaluating the value of local information and show that the mere sign of the gradient of the revenue curve at a single point can provide significant information to the seller. We then showcase how our framework can be used to run efficient price experiments. On the one hand, we develop a method to select the next price experiment that the seller should use to maximize the future robust performance. On the other hand, we demonstrate that our result allows to considerably reduce the number of price experiments needed to reach preset revenue guarantees through dynamic pricing algorithms.

3 - Active Exploration via Autoregressive Generation of Missing Data**Tiffany Cai, Columbia University, New York, NY, United States**, Hongseok Namkoong, Daniel Russo, Kelly Zhang

We pose uncertainty quantification and exploration in online decision-making as a problem of training and generation from an autoregressive sequence model, an area experiencing rapid innovation. Our approach rests on viewing uncertainty as arising from missing future outcomes that would be revealed through appropriate action choices, rather than from unobservable latent parameters of the environment. This reformulation aligns naturally with modern machine learning capabilities: we can i) train generative models through next-outcome prediction rather than fit explicit priors, ii) assess uncertainty through autoregressive generation rather than parameter sampling, and iii) adapt to new information through in-context learning rather than explicit posterior updating. To showcase these ideas, we formulate a challenging meta-bandit problem where effective performance requires leveraging unstructured prior information (like text features) while exploring judiciously to resolve key remaining uncertainties. We validate our approach through both theory and experiments. Our theory establishes a reduction, showing success at offline next-outcome prediction translates to reliable online uncertainty quantification and decision-making, even with strategically collected data. Semi-synthetic experiments show our insights bear out in a news-article recommendation task, where article text can be leveraged to minimize exploration.

4 - K-Svd with Gradient Descent**Yassir Jedra, Massachusetts Institute of Technology, Cambridge, MA, United States**, Emily Gan, Devavrat Shah

The emergence of modern compute infrastructure for iterative optimization has led to great interest in developing optimization-based approaches for a scalable computation of k-SVD, i.e., the $k \geq 1$ largest singular values and corresponding vectors of a matrix of rank $d \geq 1$. Despite lots of exciting recent works, all prior works fall short in this pursuit. Specifically, the existing results are either for the *exact-parameterized* (i.e., $k = d$) and *over-parameterized* (i.e., $k > d$) settings; or only establish local convergence guarantees; or use a step-size that requires problem-instance-specific oracle-provided information. In this work, we complete this pursuit by providing a gradient-descent method with a simple, universal rule for step-size selection (akin to pre-conditioning), that provably finds k-SVD for matrix of any rank $d \geq 1$. We establish that the gradient method with random initialization enjoys global linear convergence for any $k, d \geq 1$. Our convergence analysis reveals that the gradient method has an attractive region, and within this attractive region, the method behaves like Heron's method (a.k.a. the Babylonian method). Our analytic results about the said attractive region imply that the gradient method can be enhanced by means of Nesterov's momentum-based acceleration technique. The resulting improved convergence rates match those of rather complicated methods typically relying on Lanczos iterations or variants thereof. We provide an empirical study to validate the theoretical results.

MD29

Building B Level 2 B203

Long-Term Expansion Planning of Power Systems with integrated complex systems

Invited Session

ENRE: Electricity

Chair: Enzo Sauma, Pontificia Universidad Catolica de Chile, Santiago, Chile

1 - Modeling Aging Processes in Bess to Assess the Impact on Power System Expansion Planning**Erick Ibacache, Pontificia Universidad Catolica de Chile, Santiago, Chile**, Enzo Sauma

The global energy transition has entailed the massive introduction of variable and intermittent renewable energy generation, which has exacerbated the challenges of maintaining the security and reliability of the power supply. Lithium-ion batteries, through BESS storage systems, have emerged as complementary alternatives to renewable energy by complementing the intermittency and variability of energy generation. However, this technology entails nonlinear and nonconvex storage capacity aging processes, making it difficult to study investment decisions in power system planning and operation models. This research proposes the incorporation of a BESS system aging through a storage capacity reduction function in an iterative MILP model for Power System Expansion Planning. The aging function is defined following the work by Johnen *et al.* (2021), considering the accumulative number of cycles for a BESS asset during the modeling horizon. The nonlinearity of the aging function is managed by applying a Piecewise linear approximation. The optimization model consists of two sub models, the first one taking into consideration investment decisions, and the other searching for the optimal operation. The model was evaluated in the Chilean Power System due to the line congestion and transmission challenges faced by the country, allowing for a more comprehensive analysis of the effects of aging and investment decisions for BESS assets. The results show that ignoring aging would overestimate the investment cost of BESS systems, but not the total installed capacity, resulting in a lower investment cost due to the delay in incorporating the asset into the Power System.

2 - Long-Term Power-System Expansion Planning Model Considering Battery Energy Storage Systems and Energy Storage in the Form of Hydrogen and/or Ammonia**Enzo Sauma, Pontificia Universidad Catolica de Chile, Santiago, Chile**

The accelerated capital cost reduction of battery energy storage systems (BESS) has set a new power-systems' paradigm. Something similar is occurring with Long-term Duration Energy Storage (LDES), such as energy storage in the form of hydrogen and/or ammonia. In this work, we propose a comprehensive long-term power-system expansion planning model, formulated as a Mixed-Integer Linear Program (MILP), to minimize investment and operational costs, while considering power-system technical constraints and including 2-hour and 4-hour duration BESS, as well as energy storage in the form of hydrogen and/or ammonia. The main finding derived from our results is that it is optimal, from a least-cost viewpoint, to have a mix of on-grid and off-grid energy storage capacity.

3 - Towards Exact Time Series Aggregation for Scalable Long-Term Power System Planning

Luca Santosuosso, Graz University of Technology, Graz, Austria, Sonja Wogrin

Achieving net-zero energy systems requires the large-scale integration of renewable energy sources, advanced flexibility technologies such as energy storage systems, and sector coupling. These developments substantially increase the complexity of long-term power system expansion planning models, which must now incorporate high-resolution input time series to accurately represent the variability of power generation, demand, and related system dynamics. Time series aggregation (TSA) is widely employed to manage this complexity; however, prevailing TSA methods rely on heuristic clustering techniques that prioritize capturing the statistical features of the input data over preserving the accuracy of the optimization model's outputs. Consequently, such TSA methods often undermine the practical applicability of the resulting aggregated models. The ERC-funded NetZero-Opt project introduces a paradigm shift in TSA by proposing a novel, performance-guaranteed clustering method grounded in Dantzig's simplex algorithm for linear programming. A key feature of this method is its ability to achieve substantial dimensionality reduction while preserving optimality exactly. Applied to an illustrative economic dispatch problem, the method compresses 8,760 hourly time steps into just three representative periods, without any loss of optimality. This breakthrough redefines TSA from a heuristic preprocessing tool into an exact, performance-guaranteed model reduction technique capable of alleviating the computational burden of large-scale, complex power system optimization models. This presentation will outline both the theoretical foundation and algorithmic implementation of the proposed TSA method, highlight its implications for computational tractability in long-term expansion planning, and discuss ongoing efforts to extend its guarantees to nonconvex and stochastic problem formulations.

4 - Modeling the Joint Expansion of Hydrogen and Power Infrastructure: a European Case Study

Miguel Carrion, Universidad de Castilla - La Mancha, Toledo, Spain, Rafael Zárate-Miñano, Carlos Merino

The transition to decarbonized power systems is one of the most complex challenges facing energy planners today. In this context, integrating green hydrogen into future energy systems may be a useful means of reducing greenhouse gas emissions in processes that are difficult to electrify. This work presents a generation capacity expansion model that optimizes the deployment of hydrogen infrastructure, including electrolyzers, compressors, storage systems, and pipelines. The model explicitly accounts for the costs of pipeline installation and compression, aiming to meet both hydrogen and electricity demands. To address the uncertainties associated with this problem, a stochastic programming approach is employed. A realistic case study based on the European continental energy system is used to test the proposed formulation. The results provide power- and hydrogen-related investment decisions by country, as well as inter-country flows of electricity and hydrogen.

5 - The Electrification Race: Probing The Power Grid/Data Center/Manufacturing Nexus

Jiyong lee, The University of Texas at Austin, Austin, TX, United States, Ilias Mitrai, Erhan Kutanoglu, Michael Baldea

The simultaneous development of artificial intelligence-related infrastructure, such as data centers, and the adoption of electrification technologies in the manufacturing sector necessitate a significant expansion of the power grid over a time horizon of several decades. To achieve economic competitiveness and guarantee that the power grid can satisfy the required demand, the investments across the different sectors must occur simultaneously, i.e., they must be coordinated. We develop a capacity expansion model to simultaneously determine the optimal investment policies for power grid expansion, construction of data centers, and adoption of electrification technologies in manufacturing. We consider the case of Texas - a hub for chemical and petrochemical manufacturing as well as data centers - and analyze optimal investment policies for different objectives (cost and green house gas emissions) and sources of uncertainty.

MD30

Building B Level 2 B204

Multi-Time Scale Stochastic Optimization in Power Systems Planning and Operation

Invited Session

ENRE: Electricity

Co-Chair: Guzin Bayraksan, The Ohio State University, Columbus, OH, United States

Co-Chair: V Varagapriya, The Ohio State University, Columbus, OH, 43210, United States

1 - Bounds for Multi-Horizon Stochastic Optimization with Application to Power Generation and Transmission Expansion Planning

V Varagapriya, The Ohio State University, Columbus, OH, United States, Giovanni Micheli, Guzin Bayraksan, Francesca Maggioni

Multi-horizon stochastic optimization problems capture sequential decision-making under uncertainty across multiple time scales such as those including both strategic (long-term) and operational (short-term) scales. Due to the complexity arising from uncertainties across horizons, direct solution approaches become computationally prohibitive. We investigate computationally efficient methods to derive bounds on the optimal value of multi-horizon stochastic optimization problems, focusing on energy systems. We develop bounding techniques based on dissecting scenario trees over only strategic, only operational, and both strategic and operational scales. We show that each technique leads to a monotonic chain of lower bounds that approximate the optimal value of the original problem. Numerical results on a multi-horizon power systems generation and transmission expansion planning problem with renewable energy generation demonstrate the computational advantages of the proposed bounding methods using various dissection and recombination techniques on large-scale problems.

2 - Graph-Based Modeling and Decomposition of Multi-Scale and Hierarchical Optimization Problems

David Cole, Princeton University, Princeton, NJ, United States

Solving multi-time scale and hierarchical stochastic optimization problems can require specialized solution schemes that exploit problem structure. To capture structure in these problems, we introduce a graph-theoretic modeling framework implemented in the open-source package Plasmo.jl. This framework places variables, objectives, and constraints into nodes of a graph and places linking constraints on edges (for constraints with variables on two or more nodes). It also supports nesting graphs within other graphs to capture hierarchical relationships. These graph-based modeling capabilities provide an abstraction for modeling various stochastic programming problems in a universal way. We also introduce a generalized approach for exploiting these graph-structures using Benders decomposition and dual dynamic programming, as implemented in the package PlasmoBenders.jl. We give case studies in power systems showing how Plasmo.jl can be used for capturing problem structure, how these structures can be exploited using PlasmoBenders.jl, and how these principles extend to stochastic programming. Overall, the graph-based approach allows for capturing complex structure in large-scale optimization problems and provides a generalized abstraction for developing and implementing decomposition approaches.

3 - Market Enhancement Through the Progressive Introduction of Risk Instruments

Iman Khajepouradvani, University of Massachusetts at Amherst, Amherst, MA, United States, Golbon Zakeri, Geoffrey Pritchard, Daniel Ralph

Since uncertainties are inherent in different elements of the electricity market, such as demand, gas prices, the intermittency of renewable energies, etc. risk-averse market participants encounter unfavorable scenarios that affect both their risk exposure and their actions. This leads to an inefficient market, where the introduction of useful financial instruments can bring market closer to its optimal potential.

Building upon our previous work, we investigate to identify effective (and efficient) instruments that hedge market participants' risks and enhance the overall market structure.

4 - Climate-Resilient Nodal Power System Expansion Planning for a Realistic California Test Case

Amelia Musselman, Lawrence Livermore National Laboratory, Livermore, CA, United States

Climate change is increasingly impacting power system operations, not only through more frequent extreme weather events but also through shifts in routine weather patterns. Factors such as increased temperatures, droughts, changing wind patterns, and solar irradiance shifts can impact not only power system production and transmission but also electric load. The current power system was not designed to be resilient towards future climates. In this work, we aim to co-optimize power generation, storage, and transmission expansion in order to develop a climate-resilient system that is able to reliably meet future demands.

We analyze the impact of climate change on power systems via an adaptation of a capacity expansion planning model that seeks to minimize costs while ensuring power system resilience under a changing climate. We model the problem as a stochastic mixed-integer program, which we implement in Pyomo and solve using the stochastic programming library, mpi-sppy, along with Gurobi. We extend a synthetic but realistic, high-resolution, test case for California, the California Test System, to include parameters required for capacity expansion planning with climate scenarios. Leveraging climate data from the Energy Exascale Earth System Model, we map climate projections onto power system parameters, focusing on changes in temperature, wind speed, solar irradiance, and streamflow, which affect load and wind, solar, and hydro generation availability, respectively. We compare investment decisions based on present-day climate with future-climate and find that accounting for future climate scenarios significantly impacts generation, storage, and transmission investment decisions.

Prepared by LLNL under Contract DE-AC52-07NA27344.

5 - Methods for Multirate Scenario Generation

Harsha Gangammanavar, Southern Methodist University, Dallas, TX, United States

Decision-making under uncertainty at multiple timescales relies on an accurate representation and an ability to simulate scenarios from the underlying multi-rate stochastic process. This talk presents statistical methodologies to simulate time series data at a desired rate after observing observations at the highest rate. We present alternative approaches that involve estimating statistical models, simulating new time series using bootstrapping techniques, and subsampling. We report the results from our simulation study evaluating the empirical performance of these alternative methods.

MD31

Building B Level 2 B205

Grid Modeling, Simulation, and Resilience

Contributed Session

Contributed

Chair: Nathan Engelman Lado, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - High-Resolution Electricity Grid Modeling for Smarter Power Planning, Emissions Tracking, and System Resilience in the United States

Braden Limb, Colorado State University, Fort Collins, CO, United States, Jason Quinn

Improving the spatial and temporal resolution of electricity grid models is critical for smarter planning, operations, and analytics in an increasingly complex energy system. Traditional emissions and electricity tracking methods report data at the balancing authority (BA) level, which obscures intra-regional variation and limits the effectiveness of data-driven decision-making.

This work presents a high-resolution modeling framework designed for sub-hourly, nodal-scale electricity tracking across the contiguous United States. The model is built on real-world substation, power generation, and transmission infrastructure, enabling accurate representation of how electricity moves through the physical grid.

Using locational marginal pricing (LMP) and power flow constraints, we simulate electricity generation, transmission, and associated emissions at over 3,500 nodes. Results reveal sharp intra-BA differences in power sources and grid constraints—for example, in Texas, renewable generation in the west faces bottlenecks delivering electricity to eastern demand centers. These insights enhance emissions accounting while also supporting broader applications such as siting analysis, grid resiliency assessment, and infrastructure investment planning.

By combining optimization, machine learning, and geospatial analytics, this platform enables utilities, system operators, and policymakers to make more informed decisions using granular, real-time data. While developed with emissions tracking in mind, the model's flexible architecture supports a wide range of operational research applications. It provides a powerful tool for analyzing electricity flows, evaluating system upgrades, and designing more adaptive, efficient energy systems—advancing smarter decisions for a better energy future.

2 - Synthesizing Representative Power Flow Datasets via Constrained Diffusion Models

Milad Hoseinpour Valoujaei, University of Michigan, Ann Arbor, MI, United States, Vladimir Dvorkin

Trustworthy machine learning (ML)-based optimal power flow (OPF) solvers requires access to large-scale and high-quality datasets. Real-world power flow data are often unavailable due to security and privacy concerns, making synthetic datasets a viable and secure alternative for developing ML models in power systems. This paper proposes a diffusion model to generate synthetic power flow datasets that preserve the statistical properties of the original data and satisfy relevant power flow constraints. To reduce training complexity and exploit domain structure, we introduce a variable-decoupling strategy that leverages the weak coupling between voltage magnitudes and active power, and between voltage angles and reactive power. Furthermore, we enforce feasibility during sampling via a gradient-guidance mechanism that steers the sampling path in a direction that minimizes the constraint violation. Numerical results demonstrate the effectiveness of our approach in producing high-fidelity and statistically representative synthetic power flow data.

3 - Integrating Conductor Health into Dynamic Line Rating and Optimal Power Flow Under Uncertainty

Geon Roh, Korea Institute of Energy Technology, Naju, Korea, Republic of, Jip Kim

The integration of carbon-neutral electricity sources faces challenges due to limited transmission infrastructure. Dynamic Line Rating (DLR) offers a promising solution by maximizing the utilization of existing transmission lines based on real-time weather data. However, transitioning to DLR introduces operational risks by pushing lines to their thermal limits. Furthermore, the impact of elevated temperature operation (ETO) on conductor health and maintenance is often overlooked.

To address these challenges, this paper 1) analyzes the risk structure in the existence of DLR and wind generation (WG) forecast errors and provides insights into power system operation, 2) models maintenance costs resulting from the ETO of overhead conductors, and 3) proposes a Conductor Health-Aware Adaptive Chance-Constrained Optimal Power Flow (CHAA-CCOPF) framework that accounts for both operational risks and the long-term costs associated with ETO.

The proposed model introduces two key enhancements to the OPF. First, chance constraints are employed to address uncertainties in DLR and WG forecasts using Gaussian mixture models and copulas. Second, maintenance costs due to hourly ETO are derived from a conductor Loss of Strength model using exponential modeling. These costs are expressed as a piecewise linear function and integrated into the objective function, with the conductor temperature distribution incorporated.

The model operates within an adaptive framework to enable periodic updates. Conductor state is updated every 24 hours using real-time operational data, while the computationally complex conductor temperature-current relationship is pre-fitted using day-ahead weather forecasts for the day-ahead dispatch.

4 - Kpg 193: a Synthetic Korean Power Grid Test System for Decarbonization Studies

Geonho Song, Korea Institute of Energy Technology, Naju, Korea, Republic of, Jip Kim

This paper introduces the 193-bus synthetic Korean power grid (KPG 193), developed using open data sources to address recent challenges of the Korean power system. The KPG 193 test system serves as a valuable platform for decarbonization research, capturing Korea's low renewable energy penetration, concentrated urban energy demand, and isolated grid structure. Clustering techniques were applied to preserve key system characteristics while maintaining computational tractability and representativeness. The system includes 193 buses, 122 generators, 407 transmission lines, and incorporates temporal weather datasets.

Its feasibility was validated through Unit Commitment (UC) and AC Optimal Power Flow (ACOPF) simulations using 2022 demand and renewable generation data. This test system aims to provide a foundational framework for modeling and analyzing the Korean power grid.

5 - Large-Scale Robust and Adaptive Planning for Generation and Transmission Expansion: California Case Study

Gabriel Mantegna, Princeton University, Princeton, NJ, United States, Emil Dimanchev, Neha Patankar, Jesse Jenkins

Electric utility planners and regulators must frequently make decisions about long-lived infrastructure under deep uncertainty surrounding cost, demand, and technology availability. However, most planning practitioners use scenario analysis alone for considering uncertainty, a planning process which can frequently miss out on infrastructure decisions that offer much greater flexibility and robustness to uncertainty. Methods exist for identifying robust and adaptive decisions in an optimization framework, but they have not yet been applied to the specific context of realistic-scale generation and transmission expansion planning. In this study we develop a method for applying two-stage adaptive robust optimization to realistic-scale generation and transmission expansion planning for the state of California, and demonstrate its advantages relative to today's deterministic planning practices.

6 - The Locational Marginal Value of Storage Under Risk of Disruption

Nathan Engelman Lado, Massachusetts Institute of Technology, Cambridge, MA, United States, Austin Saragih, Saurabh Amin

Electricity storage improves the resilience of distribution networks by supplying power during outages. Previous research has characterized the locational marginal value (LMV) of storage based on its use for electricity arbitrage and reducing line losses. However, the role of storage during outages, often a key driver of storage investments, remains underexplored in the theoretical literature. We fill this gap by explicitly incorporating outage probabilities into a stylized two-stage stochastic programming model. We model a profit-maximizing load-serving entity (LSE) that operates on a radial distribution network and whose profit is determined by its electricity purchases from the grid and sales to consumers. The LSE invests in storage and controls charge/discharge decisions subject to variations in demand, prices, and the risk of disruption.

We demonstrate that the LMV of storage differs across network locations, reflecting the ability of storage to reduce line losses when placed at each location. Moreover, we show that the LMV at each location varies depending on the distributions of prices, demand, and outage probability. Additionally, the cost of load shedding and outage probabilities significantly affect the LMV of storage. Through sensitivity analysis, we find that load shedding penalties that align retailer incentives with consumer losses are essential. Without appropriate incentives, increased outage risk can paradoxically lead to reduced storage investments, leaving consumers with less electricity. Our findings highlight the importance of explicitly considering outage conditions and incentives when planning distribution infrastructure investments, particularly as the risk of extreme weather events increases.

MD32

Building B Level 2 B206

Multi-sector Modeling to Explore Energy-climate Interactions in the Electric Power Sector

Invited Session

ENRE: Energy-Climates

Chair: Ana Dyreson, Michigan Technological University, Houghton, MI, United States

Co-Chair: Kai Pan, The Hong Kong Polytechnic University, M616, Li Ka Shing Tower, Hung Hom, Hong Kong

Co-Chair: Ana Dyreson, Michigan Technological University, Houghton, MI, United States

1 - Energy and Health Nexus: a County-Level Analysis of Power Outages in North Carolina

Rodrigo Mercado Fernandez, Loyola University Chicago, Chicago, IL, United States

This study explores the impacts of power outages on the rate of hospitalizations for various physical and mental health conditions in North Carolina. Using several years worth of power outage and health data, we merged the two datasets to quantify the impacts of power outages on household health. Our objective is to provide a comprehensive analysis of the impact of power outages on the health of North Carolina residents. We also aim to identify the most vulnerable populations during power outages. This study provides important insights for policymakers and stakeholders in the energy and healthcare sectors to address the impacts of power outages on household health.

2 - Climate-Resilient Planning of the Net-Zero European Energy System

Francesco De Marco, ETH Zürich, Zurich, Switzerland, Jacob Mannhardt, Alfredo Oneto, Giovanni Sansavini

The transition to a net-zero energy system depends heavily on variable renewable energy sources such as solar, wind, and hydroelectric power. While essential for decarbonization, these technologies are sensitive to climatic conditions, introducing new vulnerabilities—especially during extreme weather events like prolonged cold spells, droughts, and *Dunkelflaute* (periods of low wind and solar output). Climate-resilient planning under future climate uncertainty is therefore critical.

This study introduces a novel framework that combines System-Informed Scenario Reduction (SISR) with stochastic optimization to design energy systems resilient to climatic stress. The approach addresses two key challenges: (1) simplifying large climate datasets while retaining critical stress characteristics, and (2) incorporating climate uncertainty directly into optimization models.

Unlike traditional meteorological scenario selection, SISR uses energy system responses—via shadow price analysis—to identify stress events. A Wasserstein-distance-based clustering method reduces 60 climate scenarios to 5 representative ones, each reflecting unique stress patterns and assigned probabilities.

Applied to a net-zero European energy model, this framework produces cost-optimal solutions that are resilient to diverse climate scenarios. Compared to conventional single-scenario planning, it achieves equivalent resilience at an average of €14.8 billion lower cost. The analysis identifies two cost-resilience trade-off regions; in the high-resilience region, the marginal cost of resilience increases fivefold, yet remains economically favorable.

Notably, onshore wind combined with hydrogen storage emerges as the most cost-effective strategy for managing multi-day energy shortages. This work demonstrates how system-informed data reduction enables scalable stochastic optimization for climate-resilient infrastructure planning.

3 - Monash Energy: the Journey to Zero Carbon Transition

Terrence W.K. Mak, Monash University, Clayton, Australia

In this presentation, I will introduce how Monash University — the Monash Energy Institute and the Faculty of Information Technology, pave the way to zero carbon transition for the Australian and global communities.

In the first part of the talk, I will introduce major energy and grid challenges in Australia, and some of the latest solutions and options used by the energy and power system industry. How the Australian Government and the Australia Energy Market Operator visioned the future energy grid of Australia will be highlighted.

In the second part of the talk, I will introduce the vision, missions, areas of innovation, and major research contributions of the Monash Energy Institute. How the Institute impact Australia and the global community will be highlighted, and major partnerships will be presented. An introduction to the Faculty of Information Technology (IT), and how the theme of Sustainable Informatics contributes to net-zero emissions will also be discussed.

Finally, I will conclude the presentation with several interdisciplinary energy projects performed by our students at Monash University.

4 - Vehicle-to-Home Integration in Solar-Equipped Homes: Cost and Emissions Impacts Across the U.S

Jiahui Chen, University of Michigan Ann Arbor, Ann Arbor, MI, United States, Parth Vaishnav

Electric vehicles (EVs) and rooftop solar photovoltaics (PV) are both key pathways to a low-carbon energy future, but both face challenges. EV load growth strains utilities and power system operators, while rooftop solar adoption is limited by high upfront costs and declining compensation for exported energy. Vehicle-to-home (V2H) technology, which allows EVs to shift household load by charging when electricity is cheap and discharging when expensive, can turn EVs into demand flexibility assets. When combined with rooftop solar, V2H increases solar self-consumption and reduces reliance on export compensation. Consequently, homeowners will likely be less affected by changes in export energy compensation policies..

This study analyzes the cost and emissions impacts of V2H in solar-equipped homes across the U.S., under current and hypothetical billing schemes with reduced solar compensation. We account for local climate and regional differences in vehicle use (based on National Household Travel Survey data) when modeling EV energy consumption. EV charging and discharging are then optimized given EV driving demand, simulated household electricity load and simulated rooftop solar output (from ResStock database). We will compile realistic electricity price data and rooftop solar compensation rules for each region based on ratebooks and the US Utility Rate Database. As the dataset will likely be helpful to the building energy modeling community, it will be readily usable hourly time series data.

5 - Improving Rural Building Stock Modeling and Electricity Demand Projections Through Data Fusion and High-Resolution Inputs

Adnan Hilal, Michigan Technological University, Houghton, MI, United States, Ana Dyreson, Josh Quinnell, Dave Bohac, Chelsea Schelly

Electrification of heating loads on a national scale has received attention, but understanding the electrification potential and barriers in the rural north still poses a challenge. This challenge stems from the complex and dynamic relation between housing envelopes, heating demand, available energy resources, and winter and summer extremes. This study aims to inform future electrification and weatherization initiatives and to improve electricity demand modeling for electric power system resilience. A major obstacle to building energy modeling is the limited availability of high-quality, localized building data in these remote regions. This presentation focuses on data fusion to characterize single-family homes in rural Michigan, Wisconsin, and Minnesota. We integrate property assessor records, satellite imagery analyzed using machine vision techniques, and a custom housing survey to develop detailed, high-resolution inputs for ResStock, an open-source residential energy modeling platform developed by NREL. We present statistical comparisons that highlight key discrepancies between national default datasets (e.g., ResStock's baseline assumptions) and localized data sources, emphasizing the importance of using high-resolution inputs for rural modeling. We also illustrate how combining these sources enables the creation of more realistic archetype homes, which can better represent regional building stock for modeling efforts. This work represents an initial and necessary step toward modeling space heating demand in the rural north to assess the building stock's electrification resilience in cold climates. This work offers a new pathway for improving energy modeling accuracy in data-sparse regions, informs future electrification and weatherization planning, and addresses electricity resilience.

6 - Heat-Related Health Risks and Backup Power Benefits Under Extreme Weather and Power Outages

Shanshan Liu, University of Michigan, Ann Arbor, MI, United States, Parth Vaishnav

The concurrence of extreme weather events and power outages can critically challenge the availability of backup power, which is essential for operating medical equipment and maintaining safe indoor temperatures for vulnerable individuals. We employed a Reduced Complexity Building Energy Model (RC-BEM) to predict experienced indoor temperatures for residents in archetypal housing types. Simulations were conducted under various combinations of weather conditions and power supply scenarios (e.g., power on and power off) at the parcel level across 42 U.S. cities. Based on established associations between temperature and mortality risk, we assessed both the relative and absolute mortality risks for individuals. We then compared the reduction in health risks between scenarios with and without electric vehicles serving as backup power. Finally, by generating nationwide census block-level maps of vulnerability and under future climate scenarios, this study offers decisionmakers with insights to coordinate and allocate resources effectively.

7 - Evaluating Resilience of Power Sector Systems Against Deeply Uncertain Weather Futures

Vincent Carag, NREL, Golden, CO, United States

The rapid evolution of the U.S. energy sector has increased the importance of investing in power systems that are resilient to a large range of future weather conditions and events. Uncertainty in the characteristics and behavior of increasingly severe weather poses risks to the future stability and operation of the power systems that we build today. This research uses future load and renewable energy resource datasets including hydropower, wind, and solar downscaled from global climate model projections as inputs to the Regional Energy Deployment System (ReEDS), NREL's flagship capacity expansion model, to simulate the evolution of the U.S. bulk power system across multiple future weather conditions. This presentation demonstrates how the data is implemented to impact grid investments and resource adequacy contributions through the Probabilistic Resource Adequacy Suite (PRAS). Results show how future weather can impact future power sector investments and operating costs and performance, providing insights into robust investment strategies under deeply uncertain future weather conditions.

MD33

Building B Level 2 B207

HAS Student Best Paper Award Session

Award Session

Health Applications Society

Chair: Osman Ozaltin, North Carolina State University, Raleigh, NC, United States

Co-Chair: Sait Tunc, Virginia Tech, Blacksburg, VA, United States

1 - The Impact of Batching Advanced Imaging Tests in Emergency Departments

Jacob Jameson, Harvard University, Cambridge, MA, United States, Soroush Saghaian, Robert Huckman, Nicole Hodgson, Joshua Baugh

Using detailed electronic health record data from two major U.S. emergency departments (EDs), we use practice variation across physicians to uncover the operational impact of discretionary batch ordering of imaging tests. Using causal inference approaches, we find that assigning an arriving patient to an ED physician who is a “batcher” versus a “sequencer” causally increases the patient's length of stay, time to disposition, and the number of imaging tests ordered. We find evidence that the impact of discretionary test batching on length of stay is heavily mediated by ordering additional imaging tests and the probability of being admitted to the hospital post-ED service. These findings suggest this ordering strategy may lead to clinical decision-making that introduces bottlenecks in patient flow. Conversely, sequencing imaging tests poses an “information gain” advantage compared to batching: the information obtained from a prior test allows for eliminating the need for ordering some future tests. Put together, our findings indicate that discretionary batch ordering may not be an optimal strategy for managing diagnostic imaging in emergency care and that interventions to reduce discretionary batch ordering may be warranted.

2 - The National Health Service Blended Payment Scheme: Incentive Issues and Optimal Reform

Sara Mohammadi, UCL School of Management, London, United Kingdom, Kenan Arifoglu, Bilal Gokpinar

The National Health Service (NHS) in England introduced the blended payment scheme for emergency care in 2019, as a mixed prospective payment model that combines fixed lump-sum payments under block contracts with variable activity-based payments under Payment by Results (PbR). Under this scheme, hospitals receive a fixed lump-sum payment at the start of the year based on their planned activity or readiness, and a variable payment at year-end that adjusts for differences between planned and actual activity, scaled by a variable rate but capped by break-glass provisions that limit extreme financial adjustment. We develop a game-theoretic model to analyze hospitals' treatment cost and planned activity decisions under the blended payment, considering information asymmetry between hospitals and regulator. Our analysis shows that while the blended scheme improves upon block contracts and PbR, it fails to fully align hospital incentives for cost reduction and readiness. We identify three

key limitations: (i) information asymmetry makes it difficult for the regulator to set the appropriate variable rate, (ii) break-glass provisions may weaken hospital incentives for readiness, and (iii) hospitals may face long-term budget deficits. To address these issues, we propose a modified blended payment that retains the fixed payment but reform the variable payment by: (i) setting the variable rate at 100%, (ii) adjusting payments based on deviations from average actual activity across hospitals rather than each hospital's own actual activity, and (iii) eliminating break-glass provisions. Our proposed payment model aligns hospital behavior with the social optimum and ensures a balanced long-term budget.

3 - Validating Counterfactual Estimations Under General Network Interference

Sadeh Shirani, Stanford University, Stanford, CA, United States, Mohsen Bayati, Yuwei Luo, William Overman, Ruoxuan Xiong

Randomized experiments are a cornerstone of data-driven decision making. Yet, their validity can be compromised by network interference, when a unit's treatment affects the outcomes of others. We argue that identifying population-level causal effects does not require recovering the full network structure; instead, it suffices to characterize how network interactions contribute to the evolution of aggregate outcomes. Building on this principle, we introduce the causal message-passing framework that investigates how outcomes evolve as the treatment propagates through the network, thereby extracting valuable information to compensate for missing network knowledge. We formalize this through state evolution equations that describe the dynamics of outcome distributions and establish a functional parallelism across treatment scenarios: while observed and counterfactual outcomes are different, both evolve under the same mathematical rules. This result yields a systematic method for counterfactual estimation and validation with consistency theoretical guarantees. To evaluate our framework, we introduce an interference gym, a suite of semisynthetic experiments spanning networks of AI agents, opinion formation in communities, and ride-sharing systems. Extensive evaluations demonstrate the robustness of our method across diverse interference structures. Finally, we delineate the theoretical limits of this approach, showing that strong temporal trends or unstable endogenous interference may undermine identification, particularly when interference patterns shift after treatment density crosses a critical threshold.

4 - Data-to-Dose: Personalized Dosing with Data Generation and Expert Feedback

Hansraj Satyam Verma, Carnegie Mellon University, Pittsburgh, PA, United States, Sridhar Tayur, Holly Wiberg, Shixiang Woody Zhu

Personalized treatment optimization is often constrained by data scarcity, especially in settings involving heterogeneous or rare patient profiles. Standard data-driven approaches struggle in these low-data regimes, limiting their ability to provide individualized and safe dosing recommendations. We introduce GenEx, a hybrid framework that integrates preference-based Bayesian optimization with a generative model that is fine-tuned using expert feedback. Expert-provided pairwise rankings guide updates to both a surrogate model and a generative component, which proposes synthetic data points in uncertain regions when the generator is consistent with expert preferences. Theoretical analysis establishes performance guarantees through sublinear regret bounds while controlling for bias introduced by synthetic augmentation. Numerical experiments simulating personalized dosing show that GenEx achieves faster convergence and improved decision quality compared to preference-only and purely generative baselines. Combining expert feedback with generative modeling can significantly improve personalized decision-making in healthcare and other data-limited contexts. This approach allows managers and clinicians to design more effective and safer individualized treatment plans while reducing the need for extensive historical data, and it offers a scalable strategy to efficiently integrate expert knowledge in practice.

MD34

Building B Level 2 B208

The Interplay Between Learning, Optimization, and Statistics

Invited Session

OPT: Machine Learning

Chair: Lijun Ding, University of California San Diego, La Jolla, CA, United States

1 - On Squared-Variable Formulations

Lijun Ding, University of California San Diego, La Jolla, CA, United States

We revisit a formulation technique for inequality constrained optimization problems that has been known for decades: the substitution of squared variables for nonnegative variables. Using this technique, inequality constraints are converted to equality constraints via the introduction of a squared-slack variable. Such formulations have the superficial advantage that inequality constraints can be dispensed with altogether. But there are clear disadvantages, not least being that first-order optimal points for the squared-variable reformulation may not correspond to first-order optimal points for the original problem, because the Lagrange multipliers may have the wrong sign. Extending previous results, this paper shows that points satisfying approximate second-order optimality conditions for the squared-variable reformulation also, under certain conditions, satisfy approximate second-order optimality conditions for the original formulation, and vice versa. Such results allow us to adapt complexity analysis of methods for equality constrained optimization to account for inequality constraints. On the algorithmic side, we examine squared-variable formulations for several interesting problem classes, including bound-constrained quadratic programming, linear programming, and nonnegative matrix factorization. We show that algorithms built on these formulations are surprisingly competitive with standard methods. For linear programming, we examine the relationship between the squared-variable approach and primal-dual interior-point methods.

2 - Robustly Learning Single Index Models

Puqian Wang, University of Wisconsin, Madison, Madison, WI, United States, Nikolaos Zarifis, Jelena Diakonikolas, Ilias Diakonikolas

Single Index Models (SIMs) represent functions formed by composing an unknown univariate nonlinear activation σ with a linear map defined by an unknown vector w . The learning task—recovering both σ and w from i.i.d. labeled examples (x, y) , where y is a noisy evaluation of the SIM—leads to a nonconvex, often nonsmooth optimization problem, even for simple activation classes. SIMs are a fundamental model in supervised learning, capturing low-dimensional structure in high-dimensional data. While the setting with zero-mean bounded-variance noise has been well studied, more realistic formulations—where labels may deviate arbitrarily from any ground-truth SIM—are substantially more challenging. In particular, more relaxed notions of error and much stronger structural assumptions about both the activation and the distribution generating the data are required for computational tractability. Most provable guarantees in this regime have emerged only recently. In this talk, I will survey these developments and present a unifying optimization-theoretic framework based on local error bounds. These bounds capture how the gradient field remains meaningfully aligned with a target solution, thus providing a geometric “signal” that enables efficient learning with first-order methods, despite nonconvexity and noise.

3 - How Neural Networks Learn Features from Data, and How to Design Better Non-Neural Models in Practice

Daniel Beaglehole, UCSD, San Diego, CA, United States, Adityanarayanan Radhakrishnan, Enric Boix-Adserà, Mikhail Belkin

Inference from tabular data, collections of continuous and categorical variables organized into matrices, is a foundation for modern technology and science. Yet, in contrast to the explosive changes in the rest of AI, the best practice for these predictive tasks has been relatively unchanged and is still primarily based on variations of Gradient Boosted Decision Trees (GBDTs). Very recently, there has been renewed interest in developing state-of-the-art methods for tabular data based on recent developments in neural networks. In our work, we leverage insights into the mechanisms for feature learning in neural networks to design xRFM, a non-neural algorithm that combines feature learning kernel machines with a tree structure to scale to essentially unlimited amounts of training data.

On the TALENT benchmark, we show that compared to \$31\$ other methods, including recently introduced tabular foundation models (TabPFNv2) and GBDTs, xRFM achieves best performance across \$100\$ regression datasets and is competitive to the best methods across \$200\$ classification datasets outperforming GBDTs. Additionally, xRFM provides interpretability natively through the Average Gradient Outer Product.

MD35

Building B Level 2 B209

Community Engagement, Resource Optimization, and New Innovations in Public Sector Operations Research

Invited Session

Public Sector OR

Chair: Carlos Morel Figueroa, Rensselaer Polytechnic Institute, 231 Congress St, Troy, NY, 12180, United States

Co-Chair: Weixiao Huang, Worcester Polytechnic Institute, Worcester, MA, 01609, United States

1 - Talk Math to Nonprofits: a Methodology to Capture Nonprofit Values into Optimization Models**Carlos Morel Figueroa, Rensselaer Polytechnic Institute, Troy, NY, United States, Weixiao Huang, Yunus Telli, Sarah Stanlick, Andrew Trapp, Jennifer Pazour**

While nonprofit organizations face challenges that can be addressed using optimization, mission-driven values complicate the use of traditional optimization-based decisions. Nonprofits must weigh these values alongside standard objectives, and at the same time understand and trust that the decisions align with community priorities. We introduce a methodology to assist optimization modelers integrate nonprofit stakeholder values into optimization models. The methodology outlines steps for identifying and incorporating multiple values, ensuring that stakeholder priorities are accurately represented. This process includes multiple stakeholder activities to collect feedback on both the interpretation of values and model outcomes.

We demonstrate this methodology with the SWAP system, a resource sharing platform piloted by Maryland-based nonprofits. SWAP facilitates resource exchanges through integer optimization. We developed five optimization models, each addressing different nonprofit values. The output of these models were presented to nonprofit stakeholders to explore how differing stakeholder priorities influence resource exchanges. Stakeholders also provided feedback on the outputs they found most useful and the values they prioritize. This case study demonstrates how stakeholder values influence decision-making in nonprofit resource exchanges and shows the practical utility of our proposed framework.

2 - Developing Optimization Models with Cognitive Systems Engineering**Emily Tucker, Clemson University, Clemson, SC, United States, Tyler O'Brien, Steven Foster, Sudeep Hegde**

One goal of applied operations research is to improve decisions in practice. For an optimization model to be useful, the underlying problem must be understood, the formulation representative, the data appropriate, and the model's recommendations understood by the decision-makers. Current approaches for developing optimization models primarily rely on past experience, rather than underlying theory, to guide decisions on how to include stakeholders in the process. In this present, we present a new framework for applied optimization modeling with stakeholders. In each iteration, modelers develop model versions based on their current mental models of the decision process; these mental models are progressively updated based on interviews with stakeholders and thematic analyses. Stakeholders' mental models may also update as the model versions and the context change. The theoretical foundation is based on methodologies from the field of Cognitive Systems Engineering that focus on understanding emergent patterns based on systems, stakeholders, and environments. We apply the framework to a case study of locating hand sanitizer stations at a large academic institution.

3 - Considerations for Modeling and Defining Effective Interventions into Sex Trafficking**Thomas Sharkey, Clemson University, Clemson, SC, United States, Vanessa Voller, Tonique Ayler, Kelle Barrick, Teresa Forliti, Joy Friedman, Daniel Lopes da Silva, Kayse Maass, Lauren Martin, Mikki Mariotti, Christine Nelson, Yongjia Song**

Interventions designed to disrupt sex trafficking and the harm that it causes often have a goal in mind. We present considerations necessary to build models of interventions into sex trafficking and measure their effectiveness. These considerations blend existing frameworks, such as the 3P paradigm and public health approaches, with survivor-centered expertise to ensure a well-rounded perspective. We then discuss how past OR work into sex trafficking fits within these considerations and how incomplete, potentially unreliable, data influences this past work.

4 - Fairness-Efficiency Symbiosis**Violet (Xinying) Chen, Stevens Institute of Technology, Hoboken, NJ, United States, David Rea**

Ensuring algorithmic decisions induce fair solutions is of critical societal importance. Much has been written about the potential tradeoff between fairness and efficiency in worst case scenarios. Simultaneously, there is growing empirical evidence that the fairness-efficiency tradeoff is often avoidable in practice. This research seeks an understanding for when fairness and efficiency can form a symbiotic relationship, i.e., when both can be improved simultaneously. Based on the distributive justice framework, we distinguish between two fairness perspectives: equity seeks outcomes proportional to recipients' merit or needs, while equality seeks the same outcomes for all. We provide simple formulas for understanding when (a) fairness will be symbiotic with efficiency, and (b) equity will be more efficient than equality and vice versa. A core component in our formulas links symbiosis patterns with the alignment among efficiency, equity and equality. Specifically, given an efficient decision, we show that making the decision more equitable can improve efficiency (i.e., a symbiotic relationship between equity and efficiency), whereas seeking better equality will almost always result in a tradeoff with efficiency.

5 - Equity in Hunger Relief**Naurin Zoha, University of Michigan, Ann Arbor, MI, United States, Julie Ivy, Tanzid Hasnain**

Food insecurity refers to a lack of access to sufficient food for an active, healthy life. In 2023, 13.5% of households in the USA experienced food insecurity. Studies show that hunger does not affect everyone equally, with persistent disparities in food insecurity levels among different demographic groups. Food banks aim to serve these populations equitably. While previous research has focused on geographic equity—ensuring food distribution aligns with regional demand—there is a growing shift toward demographic equity, which addresses disparities among groups based on race and ethnicity, age, gender, family composition etc. This study aims to develop a multi-criteria deterministic linear programming model that incorporates both geographic and demographic equity using various equity measures. It explores the impact of these measures on food bank operations while also considering effectiveness (minimizing undistributed food) and efficiency (minimizing distribution costs). The model is then applied to a real food bank network to derive optimal policies for equitable allocation across demographic groups and geographic regions.

MD36

Building B Level 2 B210

SpORts III

Invited Session

spORts

Chair: Thomas Robbins, East Carolina University, Greenville, NC, United States

1 - Scheduling An ACC Women's Basketball Season**Kevin Hutson, Furman University, Greenville, SC, United States**

The Atlantic Coast Conference expanded to eighteen teams last year with the inclusion of three new teams, two of which reside in California. This greatly increased the amount of conference travel for league teams and brought about difficult match-up scenarios like cross-country Thursday/Sunday road games against the league's top teams. In this talk we explore scheduling options for the league to balance the competing objectives of minimizing travel costs, including environmental impact, with creating a fair schedule. We also consider many of the desired constraints from the league such as venue availability, television-network-desired matchups, etc. in the model.

2 - Adding New and Entertaining sports from the Ancient Olympics to Los Angeles 2028 and Adding An Esports Olympics**Raymond Stefani, California State University Long Beach, Lake Forest, CA, United States**

Relevant and entertaining sports have been added throughout Olympic history. Beginning in 776 BC with one running event, athletics in the Ancient Olympics involved 49% of all events. Combat sports like boxing and wrestling (32%) began in 708 BC, chariot racing (11%) began in 680 BC, equestrian events (4%) began in 648 BC, while in 396 BC artistic events (4%) like acting and lyre playing were added. Although women competed in their own Heraea Games, 6 women officially won 8 Olympic equestrian and chariot events, by owning the winning horses. The 1896 marathon linked ancient with modern. Women began competing in 1900. From 1896-1936, host nations, with IOC approval, selected sports that maintained relevance and entertainment. After WW2, locally interesting non-medal demonstration sports were contested from 1948-1992. Six of those became continuing core sports: badminton, baseball, handball, judo for women, taekwondo, and tennis. Six sports were added directly to continuing status: archery, judo for men, softball, table tennis, triathlon, and volleyball. From 2004 to 2016, trying to limit sports federations to 28, baseball and softball were dropped and golf and rugby 7s replaced them. Under Olympic Agenda 2020, Tokyo 2020 added skateboarding, sport climbing, surfing, karate, baseball, and softball. Paris 2024 included skateboarding, sport climbing and surfing, which the IOC then made continuing core sports. Eclectic breakdancing was also included. The Los Angeles 2028 Olympics will add baseball, softball, lacrosse, cricket, squash, and flag football. The IOC plans an Olympic Esports Games starting in 2027 in Saudi Arabia.

3 - Brains OR Legs: How to Organize a Fair Rogaining Competition?

Dries Goossens, Ghent University, Gent, Belgium, David Van Bulck, Joonas Pääkkönen, Benjamin Jacquet

Rogaining is an orienteering running sport where participants need to decide what control points to visit and in which order. The objective is to collect the largest possible score associated with the visited controls, without violating the time limit. While extensive literature exists on the orienteering problem from a participants' point of view, there is limited understanding of how to design a rogaining contest where physical abilities do not dominate over cognitive skills in influencing the race outcome. To create these contests, we propose a heuristic bilevel optimization approach where at the upper level organizers assign scores to candidate control points, while at the lower level participants solve the classic orienteering problem. The simulation of the selected courses by the participants results in a provisional ranking which allows to evaluate the score assignment as determined by the organizer at the upper level. We apply our methodology to the 2023 World Rogaining Championships, demonstrating the necessity of thoughtful score allocation to ensure a balanced emphasis on both skills.

4 - Does the Elo Model Work for Team sports?

Thomas Robbins, East Carolina University, Greenville, NC, United States

The Elo model is a ranking model initially designed to rank chess players that has been extended to a number of individual and team sports. In this talk we review the Elo model and investigate its usefulness in short season competitive environments.

5 - Stuck in Limbo: Evaluating Post-Shot Behavior in College Basketball

Tristan Robbins, University of Florida, Orlando, FL, United States

This project introduces and explores the concept of "limbo," a critical yet often overlooked moment in college basketball when players hesitate or remain idle after a missed shot from their team. When a player is stuck in limbo, they do not crash the offensive boards and they do not retreat to defend against the opposing team's transition, instead they stand near-motionless, seemingly shot admiring. Through analyzing and evaluating game film from the University of Florida's men's basketball team, this research categorizes post-shot decisions into three outcomes: initiating offensive rebound attempts, retreating defensively, or being stuck in limbo. The main objective of this project is to quantify how each individual players' post-shot decision affects second-chance points and transition defense. Using data gathered from film analysis, regression models were constructed to evaluate the impact of these behaviors on an overall team outcome. Overall results suggest that even small reductions in players stuck in limbo on a possession-by-possession basis can significantly improve defensive performance in transition without hindering success in offensive rebounding. By assigning value to split-second decisions following missed shots, this project provides a framework that coaching staffs can use to adjust player tendencies and refine strategy. Overall, this work contributes to the broader field of basketball analytics by offering a fresh new lens on evaluating post-shot decisions.

MD37

Building B Level 2 B211

Learning and Optimization under Uncertainty: Algorithms and Applications

Invited Session

OPT: Optimization Under Uncertainty

Chair: Jiajin LI, The University of British Columbia, Vancouver, BC, Canada

1 - On the Uniform Convergence of Subdifferentials in Stochastic Optimization and Learning

Feng Ruan, Northwestern University, Evanston, IL, United States

We investigate the uniform convergence of subdifferential mappings from empirical risk to population risk in nonsmooth, nonconvex stochastic optimization. This question is key to understanding how empirical stationary points approximate population ones, yet characterizing this convergence remains a fundamental challenge. This work establishes a general reduction principle: for weakly convex stochastic objectives, we show that a uniform bound on the convergence of selected subgradients—chosen arbitrarily from subdifferential sets—yields a corresponding uniform bound on the Hausdorff distance between the subdifferentials over any open subset of the domain. This deterministic result reduces the study of subdifferential convergence to simpler vector-valued subgradient convergence. We apply this reduction to derive sharp uniform convergence rates for subdifferential mappings in stochastic convex-composite optimization, without relying on differentiability assumptions on the population risk. These guarantees clarify the landscape of nonsmooth empirical objectives and offer new insight into the geometry of optimization problems in robust statistics and related settings.

2 - Population-Proportional Preference Learning from Human Feedback

Jiawei Zhang, MIT, Cambridge, MA, United States

Conventional preference learning methods often prioritize majority opinions when aggregating preferences from multiple evaluators, resulting in policies that underrepresent minority groups. To address this limitation, we introduce a novel preference learning framework designed to align policies proportionally with the true population distribution of evaluator preferences. Our approach infers a feasible set of evaluator population shares directly from pairwise comparison data. Using these estimates, the algorithm constructs a policy that satisfies foundational axioms from social choice theory, namely monotonicity and Pareto efficiency, along with our newly introduced axioms of population-proportional representation and population-bounded strategyproofness, which existing methods fail to meet. Additionally, we propose a softmax relaxation method that smoothly balances population-proportional representation with the selection of the Condorcet winner. Finally, we validate the effectiveness and scalability of our approach through experiments on both tabular recommendation tasks and large-scale language model alignment.

3 - Gradient Descent with Adaptive Stepsize Converges (Nearly) Linearly Under Fourth-Order Growth

Liwai Jiang, Purdue University, West Lafayette, IN, United States

A prevalent belief among optimization specialists is that linear convergence of gradient descent is contingent on the function growing quadratically away from its minimizers. In this work, we argue that this belief is inaccurate. We show that gradient descent with an adaptive stepsize converges at a local (nearly) linear rate on any smooth function that merely exhibits fourth-order growth away from its minimizer. The adaptive stepsize we propose arises from an intriguing decomposition theorem: any such function admits a smooth manifold around the optimal solution—which we call the ravine—so that the function grows at least quadratically away from the ravine and has constant order growth along it. The ravine allows one to interlace many short gradient steps with a single long Polyak gradient step, which together ensure rapid convergence to the minimizer. We illustrate the theory and algorithm on the problems of matrix sensing and factorization and learning a single neuron in the overparameterized regime.

4 - Set Smoothness Unlocks Clarke Hyper-Stationarity in Bilevel Optimization

He Chen, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Jiajin LI, Anthony Man-cho So

Solving bilevel optimization (BLO) problems to global optimality is generally intractable. A common alternative is to compute a hyper-stationary point—a stationary point of the hyper-objective function formed by minimizing/maximizing the upper-level function over the lower-level solution set. However, existing approaches either yield weak notions of stationarity or rely on restrictive assumptions to ensure the smoothness of hyper-objective functions. In this paper, we remove these impractical

assumptions and show that Clarke hyper-stationarity is still computable even when the hyper-objective is nonsmooth. Our key tool is a new structural condition, called set smoothness , which captures the variational relationship between the lower-level solution set and the upper-level variable. We prove that this condition holds for a broad class of BLO problems and ensures weak convexity (resp. concavity) of pessimistic (resp. optimistic) hyper-objective functions. Building on this, we show that a zeroth-order algorithm computes approximate Clarke hyper-stationary points with a non-asymptotic convergence guarantee. To the best of our knowledge, this is the first computational guarantee for Clarke-type stationarity for nonsmooth hyper-objective functions in BLO. Our developments, especially the set smoothness property, contribute to a deeper understanding of BLO computability and may inspire applications in other fields.

MD38

Building B Level 2 B212

Structured and Hierarchical Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Afroz Jalilzadeh, The University of Arizona, Tucson, United States

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

1 - Simultaneous Learning and Optimization for Misspecified Saddle Point Problems

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

In this talk, we study a class of saddle point (SP) problems where part of the objective depends on an unknown (misspecified) parameter that must be learned through a secondary optimization task. Such problems arise in many applications, including portfolio optimization, inventory control, and power systems, where model uncertainty plays a key role. To address this challenge, we adopt a simultaneous approach that interleaves learning and optimization at each iteration, rather than solving them sequentially. This integrated framework introduces new analytical and algorithmic challenges due to the evolving nature of the learned parameter.

We extend accelerated primal-dual (APD) algorithms proposed by Hamedani and Aybat (2021) for the considered misspecified setting. Our methods are designed to efficiently handle the coupling between the two optimization tasks, with one of them enabling an adaptive backtracking line search. We establish convergence guarantees for the proposed algorithms and demonstrate their performance through numerical experiments. The results highlight the benefits of solving optimization problems jointly and demonstrate the superiority of our algorithms in effectively handling misspecified saddle point problems.

2 - Nonconvex Semi-Infinite Constrained Min-Max Optimization

Cody Melcher, The University of Arizona, Tucson, AZ, United States, Erfan Yazdandoost Hamedani, Afroz Jalilzadeh, Zeinab Alizadeh, Lindsey Hiett

In this talk, we consider a class of nonconvex min-max optimization problems with infinite constraints, motivated by applications such as adversarial robustness and safety-constrained learning. While semi-infinite programming offers a natural framework for these problems, its theoretical development in the nonconvex setting is still limited. We propose a novel inexact dynamic barrier primal-dual algorithm and show that, assuming the Lojasiewicz inequality holds for the infeasibility residual function, the method converges to an approximate KKT solution. Finally, we present several numerical experiments on robust multitask learning with task priority to demonstrate the practical performance of our approach.

3 - Solving Stochastic Quasi-Variational Inequality Problems

Zeinab Alizadeh, university of arizona, Tucson, AZ, United States, Afroz Jalilzadeh

While Variational Inequality (VI) is a well-established mathematical framework that subsumes Nash equilibrium and saddle-point problems, less is known about its extension, Quasi-Variational Inequalities (QVI). QVI allows for cases where the constraint set changes as the decision variable varies allowing for a more versatile setting. Although deterministic and stochastic QVIs have received growing attention, existing methods almost invariably require strong monotonicity, which severely restricts their practical applicability. In this talk, we relax that requirement and introduce algorithms that work under much weaker conditions, like non-strongly monotone operators, then illustrate their power through real-world applications and numerical experiments highlighting both stability and convergence.

4 - An Intuitive Algorithm for Two-Stage Distributionally Robust Optimization

Zhengsong Lu, University of Pittsburgh, Pittsburgh, PA, United States, Bo Zeng

Distributionally robust optimization (DRO) has been adopted in many practical systems (e.g., power, logistics, and healthcare) to support their design, operations, and reliability. In this talk, we take a primal perspective to analyze DRO and directly utilize its structure to develop intuitive and efficient computational algorithms. The resulting column-and-constraint generation method and its variants are simple and application-friendly, often outperforming existing approaches. Numerical results will be presented.

MD39

Building B Level 2 B213

Recent Advances in Solution Methods for Conic Optimization

Invited Session

OPT: Linear and Conic Optimization

Chair: Pouya Sampourmahani, Lehigh University, Bethlehem, PA, United States

1 - Faster High-Accuracy Algorithm for Multi-Commodity Flow

Jan van den Brand, Georgia Tech, Atlanta, GA, United States

In this talk, we will see that graph techniques can speed up algorithms for solving fractional multi-commodity flow to high accuracy.

Since the development of efficient linear program solvers in the 80s, all major improvements for solving fractional multi-commodity flows came from improvements to general linear program solvers rather than graph techniques. This phenomenon sparked research to understand why graph techniques are unlikely to help with multi-commodity flows. Evidence was given by [Ita'79 and Ding-Kyng-Zhang'22], showing that solving multi-commodity flow is equivalent to solving general linear programs, thus one would not expect graph techniques to help.

However, these equivalence reductions create sparse graphs, so improvements to multi-commodity flows on denser graphs might exist.

In this talk, we will see that one can indeed speed up multi-commodity flow algorithms on non-sparse graphs via graph techniques, demonstrating that the problem is more combinatorial in nature than previously thought. In particular, we show that k -commodity flow on an n -vertex m -edge graph can be solved in $\tilde{O}(k^{2.5}\sqrt{m}n^{\omega-1/2})$ time.

2 - Near-Optimal Path Following in the L_2 Neighborhood

Bento Natura, Columbia University, New York City, NY, United States

We investigate a variant of the trust region interior point method (TR-IPM) which was originally given by Lan, Monteiro and Tsuchiya (SIAM J. Optim. 2009). We show that the iteration complexity of this algorithm is essentially optimal in the l_2 -neighbourhood of the central path. Namely, the iteration complexity of the TR-IPM is upper bounded by the minimum number of segments of any piecewise linear curve traversing in a close l_2 -neighborhood of the central path.

This strengthens the previous bounds that were in terms of the Dikin--Todd--Stewart conditional number of the constraint matrix. In particular, it improves the universal bound given by Allamigeon et al. (SIAM J. Comput. 2025) by $O(n^{1.5} \log n)$ when in the l_2 -neighborhood. In particular, it matches the iteration complexity of any such existing IPMs which also have direct dependence on the problem dimension and the Dikin-Todd-Stewart conditional number of the constraint matrix.

Lan, Monteiro and Tsuchiya only gave a weakly polynomial algorithm for solving the trust region problem, and left it as an open question to develop a strongly polynomial algorithm. We resolve this question by presenting a new strongly polynomial algorithm that solves a parametric minimum norm problem to very high accuracy. As a result, each iteration of our TR-IPM can be implemented in strongly polynomial time. Our algorithm for the parametric problem relies on approximately computing the singular value decomposition of an associated matrix. These singular values enable us to locate the optimal parameter value in a narrow interval, where Newton's method exhibits quadratic convergence.

3 - Acceleration Meets Inverse Maintenance: Faster ℓ_{∞} -Regression

Shunhua Jiang, Columbia University, New York, NY, United States

We propose a randomized multiplicative weight update (MWU) algorithm for ℓ_{∞} regression that runs in $\tilde{O}(n^{2+1/22.5} \text{poly}(1/\epsilon))$ time when $\omega = 2 + o(1)$, improving upon the previous best $\tilde{O}(n^{2+1/18} \text{poly}(\log(1/\epsilon)))$ runtime in the low-accuracy regime. Our algorithm combines state-of-the-art inverse maintenance data structures with acceleration. In order to do so, we propose a novel acceleration scheme for MWU that exhibits stability and robustness, which are required for the efficient implementations of the inverse maintenance data structures.

We also design a faster deterministic MWU algorithm that runs in $\tilde{O}(n^{2+1/12} \text{poly}(1/\epsilon))$ time when $\omega = 2 + o(1)$, improving upon the previous best $\tilde{O}(n^{2+1/6} \text{poly}(\log(1/\epsilon)))$ runtime in the low-accuracy regime. We achieve this by showing a novel stability result that goes beyond previously known works based on interior point methods (IPMs).

Our work is the first to use acceleration and inverse maintenance together efficiently, finally making the two most important building blocks of modern structured convex optimization compatible.

4 - Cuhallar: a Gpu Accelerated Low-Rank Augmented Lagrangian Method for Large-Scale Semidefinite Programming

Jacob Aguirre, Georgia Institute of Technology, Atlanta, GA, United States, Diego Cifuentes, Vincent Guigues, Renato D C Monteiro, Victor Hugo Nascimento, Arnesh Sujanani

This paper introduces cuHALLaR, a GPU-accelerated implementation of the HALLaR method proposed in prior work for solving large-scale semidefinite programming (SDP) problems. We demonstrate how our Julia-based implementation efficiently uses GPU parallelism through optimization of simple, but key, operations, including linear maps, adjoints, and gradient evaluations. Extensive numerical experiments across three problem classes—maximum stable set, matrix completion, and phase retrieval show significant performance improvements over both CPU implementations and existing GPU-based solvers. For the largest instances, cuHALLaR achieves speedups of 30-140x on matrix completion problems, up to 135x on maximum stable set problems for Hamming graphs with 8.4 million vertices, and 15-47x on phase retrieval problems with dimensions up to 3.2 million. Our approach efficiently handles massive problems with dimensions up to $(n,m) \approx (8 \times 10^6, 3 \times 10^8)$ at 10^{-5} relative precision, solving matrix completion instances with over 8 million rows and columns in just 142 seconds. These results establish cuHALLaR as a very promising GPU-based method for solving large-scale semidefinite programs.

5 - Advances in Interior-Point Methods for Hyperbolic Programming

Mehdi Karimi, Illinois State University, Normal, IL, United States, Levent Tuncel

Hyperbolic programming is a class of convex optimization problems that has been extensively studied from a theoretical perspective, particularly in connection with the generalized Lax conjecture. Although it generalizes semidefinite programming and the hyperbolicity cone admits a well-defined self-concordant barrier (making it a natural candidate for interior-point methods), there is currently no (as far as we know) general-purpose solver available for hyperbolic programs, apart from our software package, Domain-Driven Solver (DDS). Motivated by growing interest in the practical applications of hyperbolic programming, we have made several significant improvements to DDS, enabling it to solve larger and more diverse problems. Additionally, we introduce a new library of benchmark problems designed to support the development and evaluation of hyperbolic programming solvers.

MD40

Building B Level 2 B214

Smart Resource Allocation Under Uncertainty: Cross-Industry Innovations in Service Systems

Invited Session

Service Science

Chair: Chenhao Zhou, Rutgers University, Edison, NJ, United States

Co-Chair: Tonghua Lin, Rutgers University, 605 North 5th St, Harrison, NJ, 07029, United States

1 - Modeling Liquidity and Survival: Inventory Speculation and Bankruptcy Risk Analysis

Tonghua Lin, Rutgers University, Harrison, NJ, United States, Weiwei Chen, Xiaowei Xu

This study investigates a financial inventory management problem where firms must allocate limited cash and inventory resources under market uncertainty. When market conditions are volatile, excessive inventory may erode liquidity and increase the risk of bankruptcy. We propose a stochastic approach to model firm decision-making that prioritizes survival over profit, capturing key trade-offs between inventory replenishment and liquidation. By framing bankruptcy as a critical constraint, this leads to a multi-period, multi-objective extension of the classical newsvendor problem. The study offers a new perspective on resource allocation strategies for firms facing financial and operational pressures. We further develop a stochastic optimization model based on a Markov Decision Process (MDP) framework, proving the existence of an optimal policy and validating its effectiveness through numerical experiments. The findings provide theoretical and practical insights into how firms can improve resilience through liquidity-aware planning, especially in industries vulnerable to demand and price shocks.

2 - Enhancing Dialysis Center Resilience: Strategic Integration of Contract Clinicians and Support Staff

Chenhao Zhou, Rutgers University, Edison, NJ, United States, Tonghua Lin, Xin Ding, Weiwei Chen

Facing chronic labor shortages, U.S. dialysis providers increasingly fill critical roles with agency personnel, creating *fissured* teams. *Clinicians*—registered nurses and physicians—make treatment decisions and supervise care, whereas *support staff*—licensed practical/vocational nurses (LPNs/LVNs) and patient-care technicians (PCTs)—carry out routine tasks under that supervision. Organizational learning hinges on time-in-place: clinicians who repeatedly work at the same facility accumulate tacit process knowledge, whereas agency clinicians divide their effort across multiple sites, limiting repetition and weakening local know-how.

Drawing on the knowledge-based view and learning-by-doing theory, we compile a 2015-2023 facility-year panel that links CMS staffing reports to quality and cost-efficiency metrics for 7,000s dialysis clinics. Using IV-2SLS estimation, we quantify how the share of hours worked by agency clinicians—our measure of workforce fissuring—affects facility performance. We then examine whether stability among support staff moderates any performance penalty and whether this moderation is stronger when task interdependence is high (broad delegation scope, complex case-mix).

By framing fissured staffing as an operations-design challenge grounded in time-dependent learning, the study offers actionable guidance: maintain labor flexibility at the clinician tier but anchor institutional memory through consistent support teams. The resulting rule of thumb helps healthcare and other knowledge-intensive services balance agility with the retention of critical capabilities.

3 - A Predictive Model for Hospital-Acquired Pressure Injuries (Hapi) to Enhance Healthcare Service Quality

Xinyi Huang, University of Iowa, Iowa City, IA, United States, Bang An, Weiguo Fan, Ling Tong

Hospital-acquired pressure injuries (HAPI) pose a significant risk to Intensive Care Unit (ICU) patients, with existing assessment tools like the Braden Scale often lacking accuracy. Traditional machine learning models struggle with time-variant features, leading to loss of critical information. This study introduces a novel dual-view machine learning framework that leverages electronic health records and a deep recurrent neural network to learn from both time-variant and time-invariant features separately. The proposed method achieves a prediction accuracy of 93.37% and a recall of 41.81%, with a 30% average improvement in precision over baseline models. These results demonstrate improved HAPI risk assessment, enabling earlier and more accurate identification of at-risk patients. By facilitating timely and targeted interventions, this model supports enhanced quality of care and improves the overall service patients receive in critical care settings.

4 - Privacy Shields: a Game-Theoretic Analysis of Competing Platforms' Incentives to Adopt Protection Technologies in the Presence of Data Brokers

Min Zhang, Communication University of China, Beijing, China, People's Republic of, Zhiyong Li, Hongxin Liu

Data-intensive platforms face a triple pressure of increasing consumer privacy demands, tightening regulations (e.g., the massive fines on Meta and Twitter), and market competition, urgently necessitating a balance between data utilization and privacy compliance. This study uses an analytical game-theoretic model to investigate the privacy protection technology adoption strategies of competing platforms. We find that when the unit price of consumer's recorded data by data brokers is too low (high), neither (both) competing platform adopts the privacy protection technologies; when the unit price of consumer's recorded data by data brokers is medium, competing platforms adopt asymmetric privacy protection strategies (i.e., only one platform adopts the privacy protection technology). Interestingly, when the unit price of consumer's recorded data by data brokers is medium, neither platform adopting privacy protection technology is a prisoner's dilemma situation. Finally, we find when the unit price of consumer's recorded data by data brokers is too high, both platforms adopting privacy protection technologies is a win-win result for platforms and consumers. However, when the unit price of consumer's recorded data by data brokers is medium, neither platform adopting privacy protection technology is a lose-lose result for platforms and consumers.

Supported by National Natural Science Foundation of China (No. 72472143).

5 - Effective Distribution Strategies for Food Banks: Overcoming Uncertainty in Donations and Food Waste

Xianjia Gao, University of Illinois Urbana-Champaign, Champaign, IL, United States

Food bank, a non-profit humanitarian organization, secures charitable food donations through collaborative partnerships. Food banks are pivotal in mitigating food insecurity by distributing donations to those in need. However, they face several significant challenges. First, food banks depend on uncertain donations from diverse sources such as supermarkets, farms, food manufacturers, and individuals, leading to a highly variable supply. Second, the demand from partner agencies varies, complicating the equitable distribution of food. Moreover, approximately 35% of the distributed food (equating to about 29 million pounds annually) is perishable, making the management of food storage and distribution more complex. To address these challenges, we develop a multi-period dynamic model to help food banks allocate perishable food to agencies, aiming to maximize the value of the food, achieve equitable distribution, and minimize waste. We propose a heuristic algorithm based on fluid approximation with a provable regret upper bound. Furthermore, to tackle the dynamic environment, we further refine the heuristic through reoptimization and prove that the same regret bound can be achieved. Our results demonstrate that the proposed heuristics achieve asymptotic optimality with respect to the size of the food distribution network.

MD41

Building B Level 3 B301

AI Models for Healthcare: Opportunities and Limitations

Invited Session

Health Applications Society

Chair: Kimberly Villalobos Carballo, New York University, New York, NY, United States

1 - A Multimodal Machine Learning Approach for Detection and Prevention of Excessive Treatment of Urinary Tract Infections

Carol Gao, Massachusetts Institute of Technology, Cambridge, MA, United States, Phevos Paschalidis, Dimitris Bertsimas

Urinary Tract Infections (UTIs) account for approximately 3 million visits to the emergency department annually at a cost of approximately \$3.5 billion. Despite being a common infection, challenges persist in making an accurate UTI diagnosis. Urinalysis and urine cultures are often nonspecific and positivity does not necessarily indicate a true UTI. This often leads to systematic overdiagnosis of UTI, which can lead to adverse drug events, unnecessary utilization of antimicrobials, development of organisms with antibiotic resistance, and most importantly, delay targeted interventions and prolong hospital stays due to false positivity. In this paper, we propose a multimodal clinical support tool that will give treating providers the likelihood of pre-testing (urinalysis, urine culture) UTI diagnosis. This tool integrates language data from clinician notes with tabular data extracted from the electronic health records of more than 30,000 patients at Hartford HealthCare. To our knowledge, this is the first application of language data and multimodal AI techniques to UTI prediction. Our model also differs from existing algorithms by making predictions before any tests have been ordered. This addresses overprescription and overtesting at its root cause. Our model achieves a high AUROC of 0.843, surpassing comparative studies. Appropriate choice of thresholds, furthermore, could correspond to a 51% reduction in ordered urinalysis while accurately identifying 90% of UTI positive patients.

2 - Scheduling Surgery Requests from Outdoctors and Indoctors Under Uncertainty

Serhat Gul, TED University, Ankara, Turkey, Arsham Atashi Khoei, Melih Celik

Hospitals increasingly allow surgeons who are not full-time employees (outdoctors) to rent operating rooms (ORs) to generate revenue. However, to have their surgeries scheduled, outdoctors need to compete with the full-time hospital-employed surgeons of the hospital (indoctors) for limited resources. Balancing the needs of both surgeon types while accounting for uncertainty in surgery durations creates a challenging surgery planning problem. To address this problem, we formulate a stochastic mixed-integer programming model to select outdoctor surgery requests and schedule both indocter and selected outdoctor surgeries across ORs and days within a finite planning horizon. Key performance measures include revenue from accepted outdoctor requests, and cost of patient waiting times, and that of expected OR overtime and idle time. To solve the model, we propose a problem-based scenario reduction algorithm based on loss function minimization (LFM). We solve the LFM problem using both a mixed-integer second-order cone programming model and a sub-gradient-based heuristic. As the heuristic provides near-optimal solutions within reasonable time, we adopt it for further experiments using real surgery duration data. We compare our scenario reduction algorithm against three alternatives from the literature. Additionally, we provide insights into the benefits of incorporating outdoctor surgeries into hospital surgery planning and exploring different levels of flexibility in handling these requests. Finally, we perform sensitivity analyses on various model parameters and estimate the value of the stochastic solution.

3 - Algorithmic Bias and Physician Liability

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States, Shujie Luan, Shubhanshu Singh

With the growing use of artificial intelligence (AI) in clinical decision-making, concerns about bias—manifested as differences in algorithmic accuracy across patient groups—have intensified. In response, the U.S. Centers for Medicare and Medicaid Services (CMS) has introduced a liability rule that penalizes healthcare providers who

rely on biased algorithms that result in erroneous decisions. This paper examines the impact of this anti-bias liability rule on an AI firm's development decision as well as a healthcare provider's decision to use AI. The AI firm develops an algorithm that serves two patient groups, where achieving the same level of accuracy for the disadvantaged group is more costly. The provider then decides whether and how to use AI to make treatment decisions, balancing the reduction in clinical uncertainty against the risk of incurring anti-bias liability. We find the liability rule may induce biased use of AI: The provider may underuse AI overall and disproportionately disregard AI's recommendations for disadvantaged patients. Interestingly, the effect of liability on AI use is non-monotone: as liability increases, the provider is first less likely to use AI for disadvantaged patients, but then more likely to rely on it. Furthermore, mandating equal algorithmic accuracy across patient groups may inadvertently harm all patients, in part because such mandates may lead to overusing AI for disadvantaged patients.

MD42

Building B Level 3 B302

Sanjay and Panna Mehrotra Research Excellence Award Session

Award Session

Health Applications Society

Chair: Brian Denton, University of Michigan, Ann Arbor, MI, United States

1 - Building Smarter Healthcare Systems with Intelligent Network Design

Shan Liu, University of Washington, Seattle, WA, United States

Many healthcare resource allocation challenges can be framed as geospatial network design problems facing equity-efficiency tradeoffs. This talk will feature two ongoing research using data analytics and optimization models to allocate scarce healthcare resources within a geographic network. In the first research, we minimized the turnaround time for HIV viral load testing by placing point-of-care testing machines in a hub-and-spoke network in Kenya. We designed a user-friendly decision-support tool for Kenyan health administrators, and a queueing-location-allocation model under stochastic demand, formulated as an integer program and incorporating Conditional Value at Risk. In the second research, we focused on assessing disparities in access to high-quality trauma care among sociodemographic groups in Washington state. We developed a set of geospatial and non-geospatial quality metrics and proposed an optimization model that reconfigures hospital functions to improve trauma care quality while addressing fairness concerns. The overarching goal is to ensure timely and location-appropriate access to critical services by leveraging health data and advanced analytics to inform more effective, system-level decision-making.

2 - Towards Multi-Sectoral Decision Analytics for Public Health: Overcoming Modeling Barriers Through IEOR Research and Translation

Chaitra Gopalappa, University of Massachusetts, Amherst, MA, United States

Health outcomes are shaped not only by clinical and behavioral factors but also by intersecting systemic, structural, and social determinants. As a result, there is growing consensus on the need for multi-sectoral, disease-agnostic strategies that integrate health and non-health interventions. However, designing decision-analytic models that can quantify needs and optimize interventions across sectors presents significant technical challenges—ranging from data sparsity and heterogeneity to computational scalability and policy relevance.

In this talk, I will discuss the research trajectory in my lab to address barriers at different levels: (1) large-scale modeling (using single diseases), (2) multi-disease modeling, and (3) multi-sectoral modeling. I will highlight core IEOR research innovation, including an evolving network algorithm using deep learning for developing scalable agent-based network simulations, hybrid simulation framework for joint modeling related diseases and social determinants, copula-informed graphical modeling for social determinants model parametrization, and multi-agent reinforcement learning for optimizing subgroup-specific policies under cross-group mixing. I will also highlight translational efforts in developing decision-analytic models suitable for analyses of national and global policy decisions. Finally, I will provide an overview of ongoing efforts for research, capacity building, and translation in area of multi-sectoral decision analytics.

3 - Innovative Solutions for Healthcare Operations with Scheduling and Analytics

Mehmet Begen, Ivey Business School, Western University, London, ON, Canada

Our research focuses on scheduling and the application of operations research across a variety of industries, with a particular emphasis on healthcare. We address challenges such as optimizing operating room schedules, planning home healthcare operations, designing contracts for pharmaceutical co-development, and evaluating the cost-effectiveness of kidney transplantation. Additionally, we optimize hospital cart planning, assess the efficiency of national cancer treatment centers, forecast inbound transfers of COVID-19 patients, and compare the effectiveness of diverse COVID-19 policies. I am fortunate to collaborate with exceptional colleagues, practitioners, students, and postdoctoral researchers whose dedication propels our work. In this talk, I will highlight several of these projects, share key insights and lessons learned, and outline future research directions.

MD43

Building B Level 3 B303

Healthcare Data for Operational and Economic Insights

Invited Session

MSOM: Healthcare

Chair: Jun Li, Ross School of Business, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Xinyu Liang, INSEAD, Singapore, Singapore

1 - Do Inspection Delays Lead to Quality Decline? Evidence from U.S. Nursing Homes

Wilson Lin, Santa Clara University, Santa Clara, CA, United States, Lauren Lu, Susan Lu

This research examines the impact of inspection delays on nursing home operations. Using a 8-year longitudinal dataset from 2016-2024, we find that longer interinspection times led to a quality decline in nursing homes, as evidenced by an increase in deficiency citations.

2 - Kidney Allocation Priority and Patient Outcomes: a Regression Discontinuity Analysis

Diwas KC, Emory University, Atlanta, GA, United States, Jiayi Liu

The scarcity of deceased-donor kidneys makes priority rules central to allocation decisions. This study examines how such priority influences patient behavior and transplant outcomes by exploiting an age-based cutoff in U.S. policy. Using a regression discontinuity design, we find that priority increases both the frequency and quality of organ offers, improving graft survival. However, it also reduces living-donor transplants and leads some patients to decline otherwise viable kidneys. These findings point to a tradeoff between expanding access and minimizing unintended selectivity.

3 - Operational Impact of Traveling Nurses on Surgical Teams

M. Eric Johnson, Vanderbilt University, Nashville, TN, United States, Vikram Tiwari, Maksim Yakovlev

This paper examines the operational implications of travel nurse employment in surgical settings. We analyze how variation in nurse workforce composition impacts throughput, coordination, and learning. By examining how the integration of travel nurses affects performance, this study informs the design of staffing models under workforce turnover and surgical complexity.

4 - Promoting Collaborative Care: Relative Performance-Based Payment Models for Hospitals and Post-Acute Care Providers

Hang Ren, George Mason University, Fairfax, VA, United States, Kenan Arifoglu, Tolga Tezcan

Bundled payment models are widely used in healthcare reimbursement but they were primarily designed for conditions managed by a single provider in a centralized manner. Therefore, these models do not account for the complexities of conditions requiring post-acute care (PAC) after hospitalization, resulting in weak incentives for care coordination between hospitals and PAC providers, particularly in decentralized settings. Motivated by the Comprehensive Care for Joint Replacement (CJR) payment model introduced by CMS—which holds hospitals accountable for the cost and quality of the entire episode, including PAC—we examine the effectiveness of existing payment models in decentralized care. Our analysis shows that bundled payment models and the CJR framework, especially without gainsharing agreements, do not fully align provider incentives. To address these shortcomings, we propose a payment model that explicitly links hospital and PAC provider reimbursements. Using a game-theoretical framework, we show that this model results in socially optimal provider actions. Moreover, we establish that similar efficiency gains can be achieved within the CJR framework if hospitals and PAC providers adopt structured gainsharing agreements. Numerical simulations show that implementing this approach could reduce readmission costs by 12%, generating over \$150 million in annual savings for Medicare joint replacement procedures alone.

MD44

Building B Level 3 B304

Healthcare Decision Analytics Using AI/ML

Invited Session

Decision Analysis Society

Chair: Sonza Singh, University of Massachusetts-Amherst, Coppel, TX, United States

1 - Optimizing Hpv Intervention Strategies over TIME: a Reinforcement Learning Approach Strategies: a Case Study in Hpv Intervention Planning

Shalome Hanisha Anand Tatapudi, University of South Florida, Tampa, FL, United States, Cody Palmer

The decision-making process in healthcare, especially in the field of infectious diseases, is complex and challenging due to various factors such as disease transmission mechanisms, demographic variables, population behaviors, and available interventions. Disease modeling has proven to be an effective tool for evaluating the implications of policymakers' choices and quantifying the effects of decision-making. However, there is still a need to navigate the tradeoffs between cost-effective and efficient intervention strategies, dynamically. The evolving nature of disease and population further complicates the situation, as strategies that are effective today may become less effective in the future. Therefore, it is essential to examine how present intervention decisions will impact future outcomes and the health of the population at large. In this study, we propose the development of a reinforcement learning model to determine optimal intervention strategies, focusing on a case study of the human papilloma virus (HPV). By developing an intelligible learning model, we can evaluate comprehensive interventions, such as cervical cancer screening and vaccination strategies, at different time intervals and aim to create an optimal policy that spans an extended period. This research will help validate current decisions and provide insights into the long-term impact of our choices on the population's health. Designing effective intervention strategies is crucial for reducing the prevalence of HPV-related diseases and improving overall health outcomes. By leveraging reinforcement learning, which allows systems to learn from feedback and adjust their actions accordingly, we can identify the most efficient and cost-effective approaches to HPV intervention and prevention.

2 - Applying Systems Modeling and Machine Learning Methods to Identify Hidden Figures Underlying Quality Measures

Inas Khayal, Dartmouth College, Lyme, NH, United States

Racial and ethnic disparities persist across many diseases in the U.S., including cancer, diabetes, and heart disease. These disparities are often measured using quality metrics that compare aggregate outcomes between racial groups—typically at a single time point or over fixed intervals, such as the last six months of life. Such a static, aggregated approach may obscure dynamic patterns in care delivery. Understanding these dynamic patterns more realistically elucidates underlying behavior that can be targeted for quality improvement efforts. In this presentation, we apply systems modeling and machine learning to uncover racial inequalities in end-of-life cancer care. First, we replicate traditional quality measures using aggregated metrics. Next, we introduce longitudinal signatures—constructed using the same numerator and denominator but calculated daily over the final six months of life. This method captures temporal variations in care by borrowing from engineering principles to construct richer representations of the time-varying differences in racial group experience. Using machine learning classification, we analyze both aggregate and temporal signatures to identify patterns indicative of racial disparities. Our approach highlights how dynamic modeling can reveal "hidden figures"—previously unrecognized inequalities—in care trajectories that are masked by static measures. This work demonstrates the value of integrating systems engineering and machine learning to advance equity-focused healthcare analytics and inform decision-making.

3 - An Order-Based Bayesian Network Approach for Analyzing Diagnostic Delays and Type 2 Diabetes-Related Health Outcomes in Primary Care

Yukti Kathuria, University of Massachusetts Amherst, Amherst, MA, United States, Muge Capan, Kristen Miller, William Gallagher, Elizabeth Selden

Chronic disease diagnosis and management in primary care relies on regular and timely monitoring. Additionally, the risk for diagnostic delay is not uniform across all patient subpopulations. While previous research has explored individual risk factors or used traditional regression models to assess Type 2 Diabetes (T2D) outcomes, few studies have captured diagnostic delay and the temporal and probabilistic dependencies among variables. Even fewer have done so using a structured framework that reflects the clinical sequence of events from demographics to diagnosis to post-diagnosis care. Our study addresses this gap by applying an order-based Bayesian Network (BN) to analyze how demographic factors and diagnostic delay surrogates influence T2D related health outcomes in primary care.

Using de-identified Electronic Health Record data of adult patients (N=2288) in two primary care practices over a 7-year study period, we constructed a three-layered BN informed by literature and clinician input. The model incorporated imputed, distributionally fitted data to reflect uncertainty and patient heterogeneity. Higher diagnostic lab levels were consistently associated with increased glycemic variability, while the effect of care continuity factors on outcomes varied by delay status, highlighting distinct pathways from diagnosis to post-diagnosis across patient subpopulations.

By modeling the care pathway as an ordered, probabilistic system, our approach offers a nuanced understanding of T2D progression. This enables dynamic risk assessment, helping identify subpopulations at higher risk of poor outcomes. These insights can support personalized and timely interventions to improve diabetes care across diverse patient populations.

4 - Optimizing Multidisciplinary Care: Online vs. Offline Scheduling Approaches

Sindhoora Prakash, University of Massachusetts Amherst, Amherst, MA, United States, Hari Balasubramanian, Chaitra Gopalappa

Complex care programs aim to reduce long-term healthcare expenses and promote patient well-being and self-efficacy through coordinated, lower-cost interventions delivered by multidisciplinary teams comprising both healthcare and social service professionals. These programs face significant challenges in workforce planning and resource allocation due to the non-stationary nature of both intervention arrival rates and service rates. In this talk, we examine three distinct approaches to scheduling interventions within such programs. First, we present a Mixed-Integer Linear Programming model with perfect information, serving as an idealized performance benchmark. Second, we evaluate a simple First-Come, First-Served heuristic that offers a straightforward, easily implementable baseline. Finally, we introduce an online Reinforcement Learning-based scheduling algorithm that makes adaptive, data-driven decisions in real time. We assess the efficiency and effectiveness of these methods by analyzing the distribution of daily workload across staff members and the average delay experienced by patients, offering insights into the trade-offs between optimality, responsiveness, and practical implementation in dynamic care environments.

5 - Unsupervised Learning Approaches to Detect Patterns in the Co-Occurrence of Diseases: Evidence from Medical Expenditure Panel Survey (MEPS)

Ali Jafari, University of Massachusetts Amherst, Amherst, MA, United States, Hari Balasubramanian

Understanding the relationships between diseases is one of the crucial aspects in healthcare and clinical practice. While certain associations between diseases, such as diabetes and hypertension (high blood pressure), are well established, others remain less obvious, and may remain undetected. In this study we use the Medical Expenditure Panel Survey (MEPS 2016 - 2019) to study associations between diseases. The MEPS provides comprehensive patient information, including specific

medical conditions indicated by their corresponding International Classification of Diseases, Tenth Revision (ICD-10) codes. We structured a dataset from the MEPS records, facilitating access to individual disease profiles. Subsequently, we applied the Apriori algorithm to identify various disease combinations and association rules. Moreover, we implemented a novel unsupervised learning algorithm to perform clustering, offering deeper insights into the interrelations among diverse disease categories. Using these methodologies, we aim to enhance our understanding of disease relationships, potentially contributing to more effective healthcare strategies and interventions.

6 - Decentralized Training Decentralized Execution Using a Shared Network (DTDE-SN) a New Reinforcement Learning Approach for Epidemic Modeling

Sonza Singh, University of Massachusetts, Amherst, Amherst, MA, United States, Chaitra Gopalappa

Multi-agent reinforcement learning (MARL) methods are suitable tools for modeling highly dynamic decision analyses such as intervention analyses in epidemic outbreaks across multiple jurisdictions. MARL models allow us to consider jurisdictional mixing interactions in the decision-making policies, as they evaluate decisions specific to each jurisdiction while considering the interactions between all jurisdictions. In multi-agent reinforcement learning, there are three commonly used training and execution paradigms. Centralized training with centralized execution (CTCE) converts a multi-agent problem into a single-agent problem by considering the joint state/action space of all agents as that of a single virtual agent. Centralized training with decentralized execution (CTDE) trains agents using centralized information but executes their actions independently based on their local observations. Decentralized training with decentralized execution (DTDE) trains agents independently to optimize team rewards, and each agent regards other agents as a part of the environment. We propose a new training and execution approach, where a single deep learning model (which considers the interactions between agents) is trained for all agents but each agent makes its own independent decisions. The standard DTDE approach trains a separate network for each agent. This approach is computationally complex, and this complexity grows with the number of agents. The new method, which we call decentralized training with decentralized execution using a shared network (DTDE-SN) models interactions between agents while reducing computational complexity. We used a HIV epidemic compartmental simulation model composed of multiple jurisdictions as a case study to demonstrate that the DTDE-SN method can identify optimal intervention decisions.

MD45

Building B Level 3 B305

Managing Resources in Marketplaces

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Deniz Akturk, Washington University in St. Louis, St. Louis, MO, United States

1 - Joint Pricing, Salvaging, and Inventory Replenishment Through Store Assignment

Deniz Akturk, Washington University in St. Louis, St. Louis, MO, United States

We consider a joint problem of pricing, replenishment, and salvaging across multiple stores connected to multiple warehouses. Given an initial allocation of units across warehouses and stores, the firm must determine replenishment quantities from warehouses to stores and replenishment quantities between warehouses. Additionally, the firm dynamically determines the prices charged at each store and the number of units to be sold through secondary channels at each store. Formulated as a discrete-time, non-stationary Markov Decision Process, the problem faces significant computational challenges due to the curse of dimensionality and non-linear replenishment costs. We show that this problem can be equivalently represented as an assignment problem, where stores are assigned to specific inventory and price combinations. Through this framework, we can express the problem using linear cost functions without any loss of optimality. This equivalent representation also enables a discrete mean-field approximation, which extends the standard mean-field formulation by providing discrete, feasible decisions. We prove that the discrete mean-field policy provides a sub-linear optimality gap in the number of stores and warehouses. Moreover, we also prove that the mean-field model provides a lower bound to the optimal cost, which numerical experiments indicate is attained for larger networks. The generality of our assignment-based reformulation allows for its application to broader joint inventory and pricing problems with non-linear costs. Through numerical experiments, we show that the discrete mean-field model performs near-optimally across various scenarios, including settings with a limited number of stores and settings where all stores have varying properties.

2 - Dynamic Matching Under Unbalanced Impatience

David Chen, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Zhiyuan Chen, Ming Hu, Yun Zhou

We consider an infinite-horizon dynamic matching problem in two-sided markets, where agents on one side, say, supply agents, are long-lived, and agents on the other, say, demand agents, are impatient. Agents of high and low quality arrive at the platform sequentially, and the matching reward has a supermodular structure. Our analysis reveals that in a centralized setting, prioritizing high-quality supply for high-quality demand maximizes matching rewards; conversely, low-quality demand should be paired with low-quality supply. In decentralized scenarios, we identify a welfare-maximizing equilibrium, in which high-quality demand is matched with high-quality supply first, followed by low-quality supply. Low-quality demand is matched with high-quality supply only when the number of high-quality agents is large and with low-quality supply only when the number of high-quality agents is small and the number of low-quality agents is large. Moreover, the decentralized matching process can be coordinated to achieve the centralized one by adjusting the reward allocation. These findings suggest that matching platforms should adopt differentiated matching strategies based on agent quality, balancing short-term satisfaction with long-term efficiency. Finally, we show that allowing agents to wait may or may not enhance social welfare in a decentralized setting. If the majority of rewards favor the patient supply side, waiting can be beneficial; however, if impatient demand agents receive most of the rewards, waiting could lead to diminished social welfare. This further highlights the importance of reward allocation in coordinating matching processes and managing queues effectively.

3 - Optimal Policies for Dynamic Pricing and Inventory Control with Nonparametric Censored Demands

Boxiao Chen, University of Illinois Chicago, Chicago, IL, United States, Yining Wang, Yuan Zhou

We study the classic model of joint pricing and inventory control with lost sales. The firm does not know the demand distribution a priori and needs to learn it from historical censored demand data. We develop nonparametric online learning algorithms that converge to the clairvoyant optimal policy at the fastest possible speed.

4 - An Integrated Approach to Airline Flight Network Expansion and Fleet Assignment Decisions

M. Selim Akturk, Bilkent University, Ankara, Turkey, Özge Şafak Aydın, Deniz Akturk, Alper Atamturk

The ability to match fleet capacity with growing passenger demand is crucial for airline profitability. This work develops a mathematical model that enables airlines to assess the incremental impact of each new flight request and to determine the optimal number of leased aircraft, while simultaneously building a schedule with new market routes, allowing a reliable estimate of operational costs and revenue. Given that the costs of fuel consumption and carbon emissions represent the largest components of operational expenses, the model integrates strategic network expansion decisions with operational aircraft assignment considerations. Consequently, we account for not only the seat capacity of the aircraft alternatives but also the fuel efficiency and environmental impact of the aircraft for the fleet assignment problem. Finally, by generating a new flight schedule in the proposed integrated approach, we could analyze the impact of time-dependent ticket prices for new flight requests as a function of their flight departure times.

5 - Inventory Systems with Record Inaccuracy: Transaction Errors vs. Unobservable Loss

Naveed Chehrizi, Washington University in St Louis, Saint Louis, MO, United States

We present a continuous-time Bayesian model for managing inventory in a retail setting when inventory records are inaccurate. In our formulation, the inventory level and its record have separate dynamics. The former is driven by the demand and unobservable loss process while the latter is driven by the demand and transaction error process. We analytically characterize the dynamics of the retailer's belief on the shelf inventory level and prove that the optimal restocking policy is a threshold policy. In contrast to prior conjectures, we show that transaction errors play a prominent role in increasing the uncertainty in the shelf inventory distribution. In addition to this

confounding effect, the presence of transaction errors breaks the retailer's ability to calculate this distribution by tracking a small set of sufficient statistics: it remains a function of the entire observed history. Hence, computing the optimal restocking policy, even though we obtain its exact analytical form, remains numerically intractable. We tackle this "curse of dimensionality" by dynamically adjusting the problem's underlying information filtration. This effectively projects the dynamics of the inventory level onto a manifold that is adaptively adjusting to the information revealed by the inventory record process. Adding this extra layer of adaptation, implemented by *dynamic filtration*, enables us to reduce the dimensionality of the problem by an arbitrary degree. We show that the resulting implementable policy, which is optimal for the chosen dynamic filtration, also exhibits significant robustness to model misspecification.

MD46

Building B Level 3 B306

Ron Howard Memorial Session

Award Session

Decision Analysis Society

Chair: Carl Spetzler, 2 Oak Forest Court, Portola Valley, 94028, United States

Co-Chair: Carl Spetzler, 2 Oak Forest Court, Portola Valley, 94028, United States

1 - Ronald A. Howard's Contribution to the Field of Decision Analysis

Ali Abbas, University of Southern California, Los Angeles, CA, United States

This talk will reflect on Ronald Howard's contributions to the field of Decision Analysis. The talk will also discuss how future generations may reflect on and get exposure to his work. Finally, the talk will discuss ongoing initiatives to preserve Ron Howard's legacy.

2 - Ronald A Howard's Contribution to the Field of Decision Analysis

Jeffrey Keisler, University of Massachusetts-Boston, Boston, MA, United States

This talk will reflect on Ronald Howard's contributions to the field of Decision Analysis. The talk will also discuss how future generations may reflect on and get exposure to his work. Finally, the talk will discuss ongoing initiatives to preserve Ron Howard's legacy.

3 - Ronald A Howard's Contribution to the Field of Decision analysis

Robin Dillon-Merrill, Georgetown University, Washington, DC, United States

This talk will reflect on Ronald Howard's contributions to the field of Decision Analysis. The talk will also discuss how future generations may reflect on and get exposure to his work. Finally, the talk will discuss ongoing initiatives to preserve Ron Howard's legacy.

4 - Ronald A Howard's Contribution to the Field of Decision Analysis

Carl Spetzler, Strategic Decision Group, Portola Valley, CA, United States

This talk will reflect on Ronald Howard's contributions to the field of Decision Analysis. The talk will also discuss how future generations may reflect on and get exposure to his work. Finally, the talk will discuss ongoing initiatives to preserve Ron Howard's legacy.

5 - Ronald A Howard's Contribution to the Field of Decision Analysis

MD47

Building B Level 3 B307

Network Optimization

Contributed Session

Contributed

Chair: Bingshen Lu, 4070 Clear Creek Dr, 80401

1 - Minimum Cost Nowhere-Zero Flows and Cut-Balanced Orientations

Siyue Liu, Carnegie Mellon University, Pittsburgh, PA, United States, Karthekeyan Chandrasekaran, R. Ravi

Flows and colorings are disparate concepts in graph algorithms---the former is tractable while the latter is intractable. Tutte introduced the concept of nowhere-zero flows to unify these two concepts. Jaeger showed that nowhere-zero flows are equivalent to cut-balanced orientations. Motivated by connections between nowhere-zero flows, cut-balanced orientations, Nash-Williams' well-balanced orientations, and postman problems, we study optimization versions of nowhere-zero flows and cut-balanced orientations. Given a bidirected graph with asymmetric costs on two orientations of each edge, we study the min cost nowhere-zero k -flow problem and min cost k -cut-balanced orientation problem. We show that both problems are NP-hard to approximate within any finite factor. Given the strong inapproximability result, we design bicriteria approximations for both problems. For the case of symmetric costs (where the costs of both orientations are the same for every edge), we show that the nowhere-zero k -flow problem remains NP-hard and admits a 3-approximation.

2 - Capacitated Incomplete Hub Network Design Problem with Link Congestion

Mahsa Sheikhihafshejani, Oregon State University, Corvallis, OR, United States, Hector Vergara

This work addresses an incomplete hub network design problem with capacity constraints and congestion cost for inter-hub links. We formulate the problem to identify hub locations, design the inter-hub network, and route demand flows through the network. The objective is to minimize the total cost, including fixed setup costs for hubs and links, transportation costs, and nonlinear congestion costs on the links. The problem is formulated as a mixed-integer program, where link congestion costs are approximated using a piecewise linear function. A heuristic-based method is proposed to solve the problem efficiently. The heuristic method begins by ranking nodes and links based on demands between nodes for inclusion in the network and iteratively designing the hub network based on this ranking. The solution is further improved using a variable neighborhood search method to find quality solutions. Computational experiments on two well-known benchmark datasets demonstrate that the proposed solution method can obtain high-quality solutions within a reasonable computational time. Finally, we conduct a sensitivity analysis with different parameter settings to evaluate the hub network and the impact of link congestion.

3 - Connected Dominating Sets Interdiction Problem

Haonan Zhong, Yunnan University of Finance and Economics, Kunming, China, People's Republic of, Foad Pajouh, Oleg Prokopyev

This study introduces and investigates the problem of interdicting connected dominating sets in a network via blocking edges. Given a network with weights assigned to its vertices and blocking costs associated with its edges, the decision maker aims to block a minimum cost subset of edges to increase the minimum weight of any connected dominating set in the remaining network above some desired threshold. This problem has important applications in effectively blocking the distribution of information, products and funds within adversarial social, communication, and supply networks. We explore the computational complexity of this problem and identify the collection of its feasible solutions. We then formulate this problem as a binary integer program, introduce three classes of facet-inducing inequalities for the convex hull of its feasible solutions, and develop a lazy branch-and-cut method for solving it. We also propose a combinatorial branch-and-bound algorithm which utilizes purely combinatorial bounds to implicitly enumerate the search space. The performance of our proposed methods are then compared on randomly-generated and real-life instances.

4 - Fair Network Design Problem: An Application to EV Charging Station Capacity Expansion

Nagisa Sugishita, HEC Montréal, Montréal, QC, Canada, Margarida Carvalho, Okan Arslan, Yossiri Adulyasak, Amira Dems

This study addresses the bilevel network design problem (NDP) with congestion. The upper-level decision-maker (a network designer) selects a set of arcs to add to an existing transportation network, while the lower-level decision-makers (drivers) respond by choosing routes that minimize their individual travel times, resulting in user equilibrium. In this work, we propose two novel single-level reformulations: one based on strong duality and the other based on the value function of the lower-level problem. Unlike existing approaches in the literature, which are specialized for optimizing the total travel time of all drivers, our approach is flexible and can optimize other metrics related to individual travel times or fairness. We discuss the differences between the two reformulations, as well as their computational performance on academic test instances of the NDP. We then apply our methods to the EV charging station capacity expansion problem. We define a metric, the cost of sustainability, to measure the service quality experienced by individual EV drivers, and optimize the charging station locations to improve this metric. We present the results of experiments using the road network in Quebec, including public fast EV charging stations.

5 - Multi-Tier Network Design Optimization in Physical Internet (Pi) Based Hyperconnected Urban Logistics

Praveen Muthukrishnan, Georgia Institute of Technology, Atlanta, GA, United States, Xiaoyue Liu, Benoit Montreuil

To meet the increasing demand for faster, more reliable, and sustainable logistics services in urban areas, the Physical Internet (PI) framework proposes a transformative approach to urban freight distribution. It envisions an open, hyperconnected logistics network underpinned by a hierarchical spatial structure. In this system, an urban region is discretized into unit zones -each representing clusters of customer locations - which are aggregated into local cells, and further into larger urban areas. Correspondingly, the logistics network is organized into three interconnected tiers of hubs: Access hubs (AH) serving flows between unit zones, Local hubs (LH) managing flows between local cells, and Gateway hubs (GH) coordinating flows between urban areas. This structure facilitates the consolidation of parcel flows, promotes systemic efficiency and sustainability.

This paper contributes to the development of such a system by introducing an optimization framework for the design and placement of hubs across all tiers of the urban logistics network, informed by hub capabilities and service requirements. The framework explicitly models inter-hub flows, supports robust and modular capacity planning, and incorporates real estate data (e.g., from CoStar) to identify opportunities for repurposing existing infrastructure and buildings. Further, a case study focused on Metro Atlanta demonstrates the feasibility and effectiveness of the proposed approach, highlighting its potential to support data-driven and scalable decisions for future urban logistics systems across wider geographical regions.

6 - Large-Scale Network Optimization for Haul Ramp and Road Design in Open-Pit Mining

Bingshen Lu, Colorado School of Mines, Golden, CO, United States, Felix Ayaburi, Alexandra Newman, Dinesh Mehta

Haul road optimization in open-pit mining aims to reduce fuel consumption and operational costs by improving ramp and haul road design for large haul trucks. Traditional methods rely on heuristics and often overlook the impact of terrain, slope, and truck constraints on efficiency. We model the mine surface as a 3D graph which nodes are connected based on distance and slope thresholds. Using Python and NetworkX, we design ramps using graph theoretic path-finding algorithms. Paths minimize fuel consumption under "reasonable" characteristics, e.g., distance, gradient, turn radius. This allows us to generate realistic, fuel-efficient haul routes for Komatsu 930Es and Hitachi EH5000ACs. We read data from AutoCAD mine design files to ascertain coordinates and terrain type based on data at 3 million nodes. Outputs yield interactive plots of paths with multiple origin-destination pairs; objective function weightings generate paths with different characteristics. Current results demonstrate optimized surface paths that reduce fuel usage. Future work will extend to subsurface planning, evaluating the trade-offs between material removal and long-term haulage efficiency.

MD48

Building B Level 3 B308

Bridging Data Science and Operations for Humanitarian Impact

Invited Session

Decision Analysis Society

Chair: Amir Jamali, Worcester Polytechnic Institute, 7 Schussler Rd, Worcester, MA, 01609, United States

1 - Encoding Spatiotemporal Patterns for Counter-Trafficking Intelligence

Nickolas Freeman, The University of Alabama, Tuscaloosa, AL, United States, Greg Bott, roya shomali

Adult service websites (ASWs) host advertisements for commercial sex services. Several sources suggest that some portion of advertisements can be linked to victims of human trafficking. Law enforcement, non-profit organizations, and researchers have focused on transforming ASW data into actionable intelligence for counter-trafficking interventions. However, existing analytical approaches largely rely on analyzing advertisement text, overlooking geographical and temporal movement patterns. This research examines methods to represent and analyze the spatiotemporal patterns in ASW data. We develop techniques to transform sequences of geographical locations and timestamps into vector representations that enable similarity assessment between different movement profiles. These vector-based approaches allow analysts to identify common trafficking corridors, link seemingly unrelated advertisements, and detect anomalous movement patterns. We evaluate several methods for constructing representations, compare their performance using a large multi-site dataset, and demonstrate how incorporating spatiotemporal analysis can enhance counter-trafficking efforts.

2 - Multi-Objective Humanitarian Logistics in Flood-Affected Urban Areas: Coordinated Truck-Drone Delivery with Fairness and Disinformation Considerations

Ramin Talebi Khameneh, Stevens Institute of Technology, Hoboken, NJ, United States, Jose Ramirez-Marquez, Nafiseh Ghorbani-Renani

Humanitarian logistics in post-disaster urban environments must address complex challenges, balancing rapid response with equitable aid distribution. This study addresses a multi-objective optimization problem for humanitarian aid delivery using coordinated trucks and drones, with the dual goals of minimizing total travel distance and maximizing fairness among aid recipients. We develop a mathematical model where each drone operates in tandem with a truck from a shared depot, navigating urban constraints such as damaged infrastructure and limited access. To solve this NP-hard problem, we employ both an Exact solver and the NSGA-II metaheuristic algorithm, enabling a comparative evaluation of performance and solution quality.

We validate our approach through two case studies: (1) the Hoboken flood scenario, examining variations in recipient population sizes; and (2) the West Kentucky flood, where we introduce the novel dimension of weaponized disinformation, modeled as false service requests. In this second case, we incorporate Social Vulnerability Index (SVI) data to probabilistically determine whether to serve potentially fake requests, with higher-SVI nodes receiving the benefit of the doubt to enhance social equity.

The results demonstrate that while Exact methods are suitable for small-scale instances, NSGA-II offers scalable, high-quality solutions across both objectives. These findings contribute to the development of resilient, data-driven disaster response systems capable of addressing both logistical efficiency and ethical distribution concerns under real-world uncertainties.

3 - Perception-Enhanced Vulnerability Analysis: a Human-Centric Vulnerability Metric Framework Using Multimodal Social Intelligence for Disaster Response

Amirhossein Dezhbor, Stevens Institute of Technology, Hoboken, NJ, United States, Rosario Contreras Galvan, Jose Ramirez-Marquez

Traditional flood risk assessment models rely on static vulnerability indices that fail to capture real-time community conditions during disaster events. This research develops a dynamic risk assessment framework that integrates social perception through multimodal social media analysis to enhance flood vulnerability metrics for humanitarian operations. The methodology builds upon conventional risk assessment by introducing a social perception modifier derived from real-time community data. System vulnerability encompasses population vulnerability based on age demographics and disability status, infrastructure vulnerability considering critical facility density and building age, and emergency response capacity. Flood intensity is modeled using physics-based hydrological simulation incorporating precipitation,

topography, and drainage systems. The key innovation lies in the social perception component, extracted through three-layer analysis: content analysis of text, images, and videos using natural language processing and computer vision; communication flow analysis measuring information volume and propagation patterns; and temporal alignment analysis across warning, impact, and recovery phases. The enhanced risk assessment integrates social perception as a dynamic modifier reflecting community-reported severity and spatial impact distribution. This approach transforms static vulnerability assessment into a responsive system incorporating local knowledge and situational awareness from affected populations. Implementation in Guadalajara, Mexico demonstrates the framework's effectiveness in capturing community-perceived risk variations that traditional metrics miss. The methodology addresses critical gaps in humanitarian operations by enabling real-time risk recalibration based on population feedback, supporting targeted emergency response allocation. The framework provides a replicable methodology for integrating human-centered data into quantitative risk models, advancing disaster management theory and humanitarian operations practice.

4 - Bridging Artificial Intelligence and Large Language Models for Poverty Research: Fine-Tuning Llama 3.2 1B for Multidimensional Poverty Prediction

Zahra Zarei Ardestani, Worcester Polytechnic Institute, Worcester, MA, United States, Aaron Belikoff, Raha Moraffah, Robert Krueger

The integration of artificial intelligence (AI) and large language models (LLMs) in poverty research remains unexplored. While machine learning and AI have been applied in various domains, their use in poverty analysis is limited, and no prior studies have leveraged state-of-the-art AI for this purpose. In this research, we introduce a novel approach that incorporates LLMs into poverty research by fine-tuning LLaMA to predict the multidimensional poverty index (MPI).

Traditionally, calculating MPI requires converting survey data into a structured format suitable for statistical analysis, followed by complex coding and computations. This process is labor-intensive, time-consuming, and demands expertise in statistics—resources that are often scarce in low-income countries. The financial and logistical challenges associated with these traditional methods frequently lead to delays or even the suspension of poverty analysis, hindering effective policy interventions.

To address these challenges, we propose a method that eliminates the computational overhead associated with survey analysis. By fine-tuning LLaMA 3.2 1B, we enable the model to directly interpret survey responses in text format and generate poverty index predictions without the need for numerical conversions or extensive calculations. This breakthrough significantly streamlines the analysis process, making poverty research more accessible and feasible for resource-constrained regions.

Our supervised fine-tuning approach resulted in 98% accuracy in predicting poverty indices, demonstrating that the model effectively understands the concept of poverty. This research highlights the transformative potential of AI and LLMs in humanitarian applications, paving the way for more efficient poverty assessment, and helping policy makers in addressing poverty and allocating resources.

5 - Optimizing Criminal Investigation Resource Allocation via Predict-Then-Optimize Framework

Vasuki Garg, North Carolina State University, Raleigh, NC, United States, Osman Ozaltin, Maria Mayorga, Sherrie Bosisto

Thousands of Illicit Massage Businesses (IMBs) are estimated to be operating in the United States by disguising themselves as legitimate establishments while exploiting trafficked workers, thereby harming both the victims and the massage industry. The increasing digital presence of these illicit businesses presents an opportunity for detection, a crucial task for law enforcement and social service agencies aiming to disrupt their operations. However, investigative agencies have limited resources and are unable to investigate every business, making it essential to allocate resources optimally. While state-of-the-art machine learning models can predict the likelihood of a business being illicit and help prioritize investigations, these predictions are often decoupled from downstream resource allocation decisions, such as investigation time and budget. Our research bridges this gap by leveraging business reviews from Yelp.com, enriched with data from RubMaps.ch, U.S. Census records, GIS data, and licensing information. We integrate predictive analytics with prescriptive decision-making through a predict-then-optimize approach, enabling resource-aware investigation prioritization.

MD49

Building B Level 3 B309

Emerging Topics in Operations Management

Invited Session

MSOM: Supply Chain

Chair: Lifei Sheng, University of Houston Clear Lake, Houston, TX, United States

1 - Large Language Models in Hiring Evaluation: Bias Towards Social Class Background

Yao Yao, University of Houston, Houston, TX, United States, Meng Li, Lai Wei

Large Language Models (LLMs), such as ChatGPT, with their advanced reasoning capabilities, offer significant potential as decision-support tools in hiring evaluations. LLMs have been found to excel at content generation, but their fairness and reliability as evaluators remain unknown. Prior research has documented that human recruiters exhibit bias towards candidates' social class backgrounds, but it is unclear whether, why, and when LLMs replicate or mitigate such biases. In this study, we conduct controlled experiments systematically manipulating social class signals. Both LLMs and human evaluators are tasked with assessing candidates from various social class backgrounds. Our experimental approach aims to identify the stages and types of signals that contribute to evaluative bias.

2 - Selling and Renting Mechatronics

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

Motivated by the rise of mechatronics (a fusion of digital control systems and physical products, such as automobiles where software can unlock advanced functionalities), we examine how mechatronic firms should market their products. Employing a Stackelberg game model, we investigate two strategies: selling both high- and low-end products or offering just the low-end product with an option for consumers to rent the high-end functionality. Our analysis begins by characterizing consumer behavior, followed by deriving the optimal selling and renting strategies. We identify a critical physical-vs-digital threshold: below it, selling is more advantageous, while above it, renting prevails. Our findings reveal that the optimal strategies of mechatronic firms may resemble those of either traditional physical goods firms or digital goods firms. However, merely possessing mechatronic capabilities alone does not guarantee the successful implementation of digital-like strategies. We also assess the impact on consumer welfare and discover that below the physical-vs-digital threshold, both the firm and consumers benefit from the selling strategy, resulting in a win-win situation. However, above the threshold, a misalignment may emerge between the firm's strategy and consumer welfare. Curiously, this misalignment typically occurs when the firm offers its low-end product for free while charging a high price for renting the high-end functionality, capturing almost the entire consumer surplus. This research advances the literature on business models traditionally available to physical and digital goods firms separately, explaining how mechatronic firms can merge strategies from the physical and digital realms, leveraging digital technologies to enhance product offerings.

3 - Rethinking Incentive Design in Response to AI Adoption

BO PANG, National University of Singapore, Singapore, Singapore, YUTING ZHU

The widespread adoption of AI across industries is fundamentally reshaping workplace dynamics and raises critical questions about the design of incentive schemes. This paper addresses a central question: how does the integration of AI systems influence the structure of organizational incentives? To explore this, we develop a game-theoretic model within a contract-theoretic framework that captures the interaction between AI systems and human agents. Our model examines how the presence of AI affects both individual incentives to exert effort and the composition of the workforce in terms of skill—capturing what we define as the incentive effect and the selection effect, respectively. By characterizing distinct patterns of human-AI interaction, the model uncovers the underlying mechanisms driving each effect. Our findings offer new insights into how the interplay between AI and human agents shapes optimal incentive design across different organizational contexts.

4 - Data-Driven Joint Flexibility and Capacity Planning Under Service Level Considerations

Min Liu, Tsinghua University, Beijing, China, People's Republic of, Shixin Wang, Yong Liang

We study a joint flexibility and capacity planning problem in a resource allocation system with multiple resources and demands. The decision-maker aims to ensure a high probability of meeting demands with varying service-level requirements. We propose a two-stage distributionally robust chance-constrained framework to model the problem, addressing the challenge of demand data scarcity, and present a solution approach based on a reformulation of the problem as a mixed-integer linear programming (MILP) problem. Leveraging our model and analysis, we investigate the performance of the widely celebrated long-chain structure under service-level constraints, where resource capacities are treated as decision variables and may vary across resources. Our findings complement the existing literature, which predominantly examines the performance of chaining structures under the maximum flow objective with fixed identical resources. Specifically, we identify conditions under which the long-chain structure is optimal and those under which it is suboptimal. Our findings suggest that low demand volatility and restriction to identical capacities are strong indicators of the long-chain structure being optimal. However, allowing for a high tolerance of unmet service-level requirements or introducing significant demand volatility can make the long-chain structure suboptimal even under identical capacities. Additionally, allowing for deviations from identical capacities typically leads to the suboptimality of the long-chain structure. Interestingly, the long-chain structure can maintain its optimality in situations with low demand volatility, even without the restriction of identical capacities.

5 - Managing Road Safety in Gig Work: Can Recommender Systems Save Lives?

Varun Karamshetty, National University of Singapore, Singapore, Singapore, Nakyung Kyung

Food-delivery drivers make frequent, complex decisions about which tasks to accept, often while on the move. This cognitive load can lead to distraction and increase the risk of accidents. Recommender systems, which streamline task selection by offering optimized suggestions, may help mitigate this risk. We investigate whether adopting a task recommender system reduces accident rates among food delivery drivers. We also examine what happens after adoption—specifically, whether drivers maintain adherence to recommendations after experiencing road accidents. Traumatic events such as accidents can prompt individuals to reclaim a sense of personal agency. Because recommender systems may be perceived as undermining that agency, we explore whether accidents lead to reduced system adherence. We partnered with South Korea's largest food-delivery platform and analyzed behavioral and safety data from 8,883 drivers, some of whom adopted a task recommender system. Using a quasi-experimental design and a Difference-in-Differences framework combined with Propensity Score Matching, we compared outcomes between adopters (treatment) and matched non-adopters (control). Drivers who adopted the system experienced an 85% reduction in at-fault accidents and a 65% reduction in total accidents. They also saw a 9.52% increase in earnings and a 0.84 percentage point increase in on-time deliveries. However, within the adopters, adherence to recommendations declined by 2.9% following an accident. Our study provides the first empirical evidence on how adopting a task recommender system impacts the physical safety of gig workers, and how road accidents impact the post-adoption recommendation adherence.

MD50

Building B Level 3 B310

Mostly Queueing: Abandonment, Control, Pricing, and Signaling in Stochastic Systems

Invited Session

MSOM: Service Operations

Chair: Philipp Afèche, University of Toronto, Rotman School of Management, Toronto, ON, Canada

1 - Quantitative Gaussian Approximation for Queues with Abandonment via Wasserstein Bounds

Zedong Wang, ISyE Georgia Tech, Atlanta, GA, United States, Siva Theja Maguluri

The steady-state distribution of $M/M/1+M$ queue—a single-server queue with exponential rate of abandonment—has been shown to converge to Gaussian in heavily overloaded regime as the abandonment rate goes to zero. However, for pre-limit abandonment rates, the distributional behavior remains less understood. In this work, we consider a new tractable framework, well suited for Birth Death Process, to quantitatively characterize the steady-state queue length by establishing upper bounds on the **Wasserstein- p** distance between the normalized queue length and a standard normal distribution. This framework not only recovers classical Gaussian convergence but also yields Berry–Esseen-type convergence rate. Furthermore, the tightness of our bound captures subexponential and Bernstein-type tail behavior, offering a better understanding of the pre-limit queue behavior in heavily overloaded regimes.

2 - An Online Learning Approach to Pricing in Double-Sided Queues

Zicheng Wang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Saif Benjaafar, Xinyun Chen, Xiaotang Yang

We study pricing in a double-sided queue with impatient agents (customers on one side and servers on the other). The service provider seeks to set prices to charge customers and wages to pay servers to maximize cumulative expected profit. We propose an online learning framework for settings where the demand and supply functions, as well as the abandonment rates, are unknown. The algorithm employs a user-level randomization scheme to estimate the gradient of the objective function. We validate the effectiveness of the algorithm through both theoretical analysis and simulation experiments.

3 - An Equilibrium Solver for A Dynamic Queueing Game

Denise Cerna, University of California, Berkeley, Berkeley, CA, United States, Chiwei Yan, Hongyao Ma

Consider the allocation of deceased donor organs to patients in transplant wait lists, or the dispatch of ridesharing trips to drivers lining up at airports. In these dynamic queueing games, heterogeneity in items and agents' ability to decline offers often induce substantial cherrypicking, resulting in high rates of unfulfilled trips and discarded organs. Existing research typically focuses on simple, easy-to-analyze dispatch rules, sometimes making fluid assumptions for tractability. In this work, we introduce an analytical, best-response-based solver for general dispatch rules with completely closed-form updates. It computes agents' equilibrium strategy regarding whether to accept their received items and whether to join the queue or renege given the evolving queue length and their queue positions. Assuming homogeneous agents and stationary Poisson arrivals, we first provide a necessary condition for a Markov Perfect Equilibrium. We also prove that this equilibrium exists and that our solver always converges for dispatch rules where items are offered monotonically down the queue, and dispatch probabilities over queue positions are independent of queue length. Via extensive simulations, we show that the solver recovers known results for policies that can be analyzed theoretically, and provides insights beyond what had been gleaned from previous analysis. Moreover, the solver converges even for policies without theoretical guarantees, allowing market designers to gain insight into more complex policies better suited for practice.

4 - Dynamic Control of a Make-to-Order Manufacturing System Under Throughput Time Constraints: An Effective Computational Method in the High-Dimensional Case

Chun Li, Booth School of Business, The University of Chicago, Chicago, IL, United States, Baris Ata, NIAN SI

We consider dynamic control of a make-to-order production system that serves price- and delay-sensitive customers. The production system is modeled as a multiclass queueing network where the customers' delay sensitivity is modeled via throughput-time constraints. To maximize her profit while meeting these throughput time constraints, the system manager uses three different types of control: i) dynamic pricing decisions; ii) dynamic order-of-service (sequencing) decisions; and iii) outsourcing decisions upon arrival of new orders. We approximate this problem by a multi-dimensional Brownian control problem in heavy traffic, and develop a neural network-based computational method tailored for high-dimensional problems. By interpreting the solution in the context of the original model, we propose an integrated dynamic pricing, outsourcing, and scheduling policy. The effectiveness of this policy is demonstrated through a simulation study.

5 - Decentralized Signaling Mechanisms

Krishnamurthy Iyer, University of Minnesota, Minneapolis, MN, United States, Niloufar Mirzavand Boroujeni, William Cooper

We study a system composed of multiple distinct service locations that aims to convince customers to join the system by providing information to customers. We cast the system's information design problem in the framework of Bayesian persuasion and describe centralized and decentralized signaling. We provide efficient methods for computing the system's optimal centralized and decentralized signaling mechanisms and derive a performance guarantee for decentralized signaling when the locations' states are independent. This tight guarantee states that the probability that a customer joins under optimal decentralized signaling is bounded below by the product of a

strictly positive constant and the probability that a customer joins under optimal centralized signaling. We also consider more-general objectives for the system and establish that the same guarantee continues to hold. We extend our analysis to systems where the locations' states are correlated, and again derive a performance guarantee for decentralized signaling. For the correlated setting, we prove that the guarantee's asymptotic dependence upon the number of locations cannot be substantially improved. A comparison of our guarantees for independent locations and for correlated locations reveals the influence of dependence on the performance of decentralized signaling.

MD51

Building B Level 3 B311

Data-Driven Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Lennart Baardman, University of Michigan, Ann Arbor, MI, United States

1 - Assortment and Fulfillment Optimization in Online Grocery Retail

Hyun-Soo Ahn, University of Michigan, Ann Arbor, MI, United States, Lennart Baardman, Renjun Hu

Despite the rapid growth of online grocery retail in recent years, most supermarkets have not turned this growth into a profitable business model. One of the key reasons is the excessively high picking and delivery cost in an industry with low profit margins. Technological innovation, in the form of the Micro Fulfillment Center (MFC), has great potential to reduce costs by improving the efficiency in picking operations. The MFC is a small local warehouse, often attached to a brick-and-mortar store or regular fulfillment center, which very efficiently fulfills a subset of items that are available in the assortment of the online store. For a grocery retailer with an online store and an MFC, it is important to decide which items are offered in the assortment of the online store and which items are fulfilled from the MFC. This is especially important in online grocery retail where cart abandonment is ubiquitous. If customers cannot find the items they are looking for, they can abandon their shopping cart and move to another online grocery retailer. In this work, we formulate the assortment and fulfillment problem for an online grocery retailer with an MFC whose customers exhibit cart abandonment behavior. We show that this problem is difficult to solve, and hence, we develop an algorithm to quickly compute near-optimal decisions at large scale. Our algorithm is shown to be a good approximation with theoretical results and computational experiments.

2 - How Not to Overpackage: AI for Sustainability in Hellofresh's Service Supply Chain

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

Overpackaging is a common challenge faced by meal kit services. For each box of meal kits delivered to customers, meal kit service providers must decide the type of liner materials and the quantity of freezer packs to place inside. These materials ensure the quality of the meal kits delivered; on the other hand, overpackaging leaves a large carbon footprint and imposes psychological burdens on many customers. As the world's leading meal kit company and integrated food solutions group, HelloFresh is committed to furthering the sustainable nature of its supply chain and decreasing its environmental footprint. One important step is to reduce packaging. While HelloFresh has an in-house thermal engineering team that builds the packaging guidelines with a large number of scientific experiments, artificial intelligence offers numerous possibilities for improvement. This paper introduces the Contextual Packaging problem, and investigates artificial intelligence solutions to mitigate overpackaging efficiently. We develop a contextual bandit algorithm that takes into account the weather, the conditions of transit, the box contents and other contextual information to adaptively make the packaging decision for each box. To accelerate online learning, we identify a number of structural properties of the contextual packaging problem and extract structural information from the abundantly available past delivery data. Our algorithm achieves theoretically optimal performance with an $\tilde{O}(\sqrt{T})$ regret bound where T is the horizon length. We test our algorithm on real delivery data from HelloFresh and identify points of overpackaging in the current guidelines.

3 - A Transformer-Based Approach to Choice Modeling: Capturing Substitution and Complementarity Effects

Salvador K. Dzimah, MIT Operations Research Center, Cambridge, MA, United States, Manuel Moran-Pelaez, Georgia Perakis

Many demand estimation methods tend to not consider both substitution and complementarity effects among items. The few methods that consider both effects do not address the problem of explicitly measuring each item's impact on the others' demands. This motivated us to propose a novel Transformer-based choice model to capture both effects. Our model integrates text, image and categorical product features. Its architecture enforces an additive decomposition of each item's sales into a baseline plus contributions from other items. Our proposed model captures rich interactions while being interpretable due to its additive structure. Furthermore, the model generalizes to new products. We tested the model on real-world data. It successfully detects the underlying cross-item effects and achieves a 10% reduction in wMAPE compared to alternative methods that ignore these effects. The additive structure allows us to not only detect complements and substitutes, but also to measure the effects quantitatively and provide actionable insights for retailers.

4 - Managing Product Availability in Rental Subscriptions

Lennart Baardman, University of Michigan, Ann Arbor, MI, United States, Xuening Wang

For rental subscription services, managing product availability is a crucial operational challenge. When products are frequently unavailable, subscribers face a limited assortment that lacks variety and leads many to churn. We analyze data from a clothing rental subscription service to better understand customer attrition. Using machine learning, we find that an important contributor to churn is the customer's perception of limited variety in available clothing styles. One way to address this challenge is through better inventory management. In current practice, styles are shown to customers only when fulfillment can be guaranteed. In contrast, we develop inventory management policies that balance the trade-off between style display and order fulfillment. These policies aim to improve product availability while maintaining reliable delivery. We demonstrate that our approach outperforms the existing practice, both through theoretical bounds and through numerical simulations.

MD52

Building B Level 3 B312

Powering the Future: Innovations in Electric Vehicles, Energy Storage, and Renewable Generation

Invited Session

MSOM: Sustainable Operations

Chair: Yangfang Helen Zhou, Singapore Management University, Singapore, 228492, Singapore

1 - Impact of Energy Storage on Electricity Prices and Carbon Emissions

JIN-YANG HU, Duke University, Durham, NC, United States, Kevin Shang, Can Zhang

Energy storage serves as a practical solution to address the challenges arising from the high penetration of renewable energy sources in the electricity system. In this study, we analyze the role of energy storage, its impact on prices and emissions, and provide insights with significant policy implications.

2 - Battery Operations in Electricity Markets: Strategic Behavior and Distortions

Jerry Anunrojwong, University of Washington, Seattle, WA, United States, Santiago Balseiro, Omar Besbes, Bolun Xu

Electric power systems are evolving to accommodate intermittent renewable sources and storage technologies like batteries. As privately-owned batteries gain prominence, key questions arise: How do they behave in electricity markets, and how does their strategic behavior differ from that of centrally operated batteries? We propose an analytically tractable model that captures salient features of the highly complex electricity market. We derive in closed form the resulting battery behavior and generation cost in three operating regimes: (i) no battery, (ii) centralized battery, and (iii) decentralized profit-maximizing battery. We identify three distortions in decentralized battery behavior: (1) quantity withholding, i.e., discharging less than centrally optimal; (2) shift from day-ahead to real-time, i.e., postponing some discharge

from day-ahead to real-time; (3) reduction in real-time responsiveness, i.e., discharging less in response to smoothing real-time demand than centrally optimal. We quantify efficiency loss using the Price of Anarchy metric and show that the cost of strategic behavior is always bounded between $\$9/8\$$ and $\$4/3\$$. That is, incentive misalignment always exists, but it is bounded even in the worst case. We calibrate our model to real data from Los Angeles and Houston. Lastly, we show that competition is effective at reducing distortions, but many market power mitigation mechanisms backfire, leading to higher total cost. This work offers a framework for analyzing and quantifying market power in battery operations. It underscores that while battery market power does pose a concern, its negative impact is limited compared to the overall system benefits of sufficient battery capacity.

3 - Aggregator Business Model for Vehicle-to-Grid Services

Yangfang Helen Zhou, Singapore Management University, Singapore, Singapore, Meichun Lin, Owen Wu, John Pang

The rapid growth of electric vehicles (EVs) creates a major opportunity for vehicle-to-grid (V2G) services, where EVs discharge electricity back to the grid during peak demand. Since an average vehicle is idle over 95% of the time, EVs can potentially act as mobile storage to reduce reliance on costly, high-emission peak power plants. We propose a business model where EVs participate in V2G only a few times per year—aligned with peak demand days—ensuring minimal battery wear and preserving warranties from EV manufacturers. Even with such limited use, the collective impact from a growing EV base can be substantial. An aggregator coordinates drivers, participates in the electricity market, and compensates drivers while retaining profits or maximizing social welfare.

We build a two-stage stochastic model. In the first stage, the aggregator offers a reward menu based on the number of V2G services, and then the drivers choose their participation level based on their inconvenience costs. In the second stage, the aggregator solves a stochastic dynamic program to determine the number of V2G service reserves to use in each period, given the electricity demand and the remaining reserve.

We characterise the optimal reward menu design and show that it can vary non-monotonically as the number of services increases. Using real electricity price and demand data, we demonstrate the model's economic and environmental benefits, benchmarking with common approaches.

4 - Optimal Charging Decisions for Battery Swap Stations Under Demand Charge

Zimeng Li, Zhejiang University & City University of Hong Kong, Hangzhou, China, People's Republic of, Zhankun Sun, Quan Yuan

Battery swap stations mitigate electric vehicle range anxiety by enabling rapid replacement of depleted batteries. Our study investigates the optimal charging decisions for a battery swap station utilizing both grid and renewable energy. Our model incorporates stochastic customer arrivals, uncertain renewable generation, and time-varying electricity prices. The objective is to minimize total electricity bill (cost based on energy consumption and peak demand charges) and customer waiting cost by optimizing charging decisions. We formulate this as a finite-horizon dynamic programming problem and prove that the optimal charging policy is of threshold type. We find the optimal charging quantity is nonlinear with respect to the current demand charge. To tackle the computational complexity of the high-dimensional model, we propose a deep reinforcement learning method, informed by the threshold structure, to efficiently compute the optimal charging policy.

MD53

Building B Level 3 B313

Dynamic Decision Making under Uncertainty: Models, Signals, and Experiments

Invited Session

Revenue Management

Chair: Zhe Liu, Imperial College London, London, United Kingdom

1 - Reformulations of Approximate Linear Programs for Finite-Horizon Markov Decision Processes

Jiannan Ke, Wuhan University, Wuhan, China, People's Republic of, Dan Zhang, Huan Zheng

We establish reformulations of approximate linear programs (ALPs) for a class of finite-horizon Markov decision processes (MDPs). These reformulations are either equivalent formulations or relaxations of the original ALPs. Our main result is established via constraint reformulations and duality arguments. In particular, it allows us to find equivalent reformulations of ALPs by checking a strong duality condition. We apply the framework to network revenue management and the finite-horizon restless bandit problem. For network revenue management with and without customer choice, we provide new, simplified proofs for the main results in Vossen and Zhang (2015) under the usual assumption of a binary consumption matrix. When the consumption matrix is non-binary, we establish new upper bounds under both affine and SPL approximations. For the finite-horizon restless bandit problem, we establish an equivalent reformulation under the assumption of a unit cost for pulling each arm and a relaxed upper bound when the cost is more general. A numerical study shows that the quality of the bounds and heuristic policies of the reformulations are competitive with alternative solution methods but at substantially reduced computational cost. Our work applies to a broad class of finite-horizon MDPs and has potential applications to many practical problems.

2 - Dynamic Persuasion Strategies for Mitigating the Spread of Fake Content

Wenjuan Li, HKUST Business School, Hong Kong, China, People's Republic of, Dongwook Shin

We examine how digital content platforms can design signaling mechanisms to shape consumers' engagement with potentially fake content. Initially, the platform and consumers face uncertainty about content veracity, but the platform can conduct random inspections to verify it. Based on inspection results, the platform strategically sends signals to consumers to ultimately maximize revenue. We analyze the structural properties of optimal public and private policies. Our findings show that both policies disclose inspection results with strategic delays, which enhances platform revenue but also increases consumers' exposure to fake content. Furthermore, unlike the optimal public policy, we find that the optimal private policy selectively discloses the inspection results to some consumers; consequently, the optimal private policy generates higher revenue than the optimal public policy but significantly undermines consumer welfare. We explore various strategies to reduce the spread of fake content while ensuring profitability. We find that enhancing the platform's inspection capability boosts the platform's profits but fails to curb the spread of fake content. Alternatively, our analysis suggests that implementing subscriptions and government penalties can effectively reduce fake content while enhancing the platform's revenue.

3 - Dynamic Decision Making with Operational Data Analytics

J Shanthikumar, Purdue University, WEST LAFAYETTE, IN, United States, Qi Feng, Yang Yang

We consider dynamic decision making with limited knowledge of the reward distribution. Under a parametric setting we show that a uniformly optimal data integrated dynamic policy can be derived for some classes of inventory and revenue management problems. We will use a simple revenue management problem to illustrate the basic idea.

4 - Optimal Resource Allocation in the Experiment-Rich Regime

Fenghua Yang, University of Michigan, Ann Arbor, MI, United States, Stefanus Jasin, Dae Woong Ham

Motivated by the large-scale experimentation practices of modern companies, we examine the problem of optimally allocating a fixed N number of customers across M concurrently running A/B testing experiments. While previous literature focuses on optimal allocation under a minimax framework minimizing mean squared error (MSE), we extend previous results under a new hypothesis testing framework that is also managerially relevant for decision making. Specifically, instead of minimizing MSE, we minimize type-2 error while always maintaining a valid type-1 error control. We provide optimal allocation results under two distinct cases: known variances and unknown variances. We find that minimizing MSE does not necessarily minimize type-2 error, especially under the case when the gap of minimum detectable effect is large. Extending our analysis to objectives beyond minimax, we show that, once the total sample size is sufficiently large, the asymptotically optimal allocations under the weighted-minimax and weighted-average criteria share the similar structure.

MD54

Building B Level 2 B216

From Data to Decisions: AI in Revenue and Healthcare

Invited Session

Revenue Management

Chair: Mingxi Zhu, Georgia Institute of Technology, Atlanta, GA, 30033, United States

1 - Data-Driven Contextual Pricing with Semi-Parametric Models**Xiangyin Chen, Cornell University, Ithaca, NY, United States, Li Chen, Meng Qi**

We introduce a unified framework for the data-driven revenue-maximizing contextual pricing problem. Along with available contextual features, we consider two types of data: (1) observed customer valuations (also referred to as willingness-to-pay) and (2) customers' binary purchase data and posted prices, without knowledge of the prior pricing policy. We assume a semi-parametric model of customer valuation with a log-linear function of contextual features and a distribution-free noise term. This model enables an elegant solution for the optimal pricing policy that separates the dependencies of the contextual features and the noise term. Leveraging this solution structure allows us to develop efficient algorithms for the above two data settings. We show asymptotic optimality and finite-sample performance guarantees for these algorithms under general, estimator-agnostic conditions. Finally, we conduct comprehensive numerical experiments with both synthetic and real data. Tested on synthetic data, our algorithms demonstrate superior performance with various underlying distributions of the noise term. Applied to a real-world online auto loan dataset, our algorithms outperform commonly adopted state-of-the-art benchmarks.

2 - Market Power OR Data Power: when Do Competitors Choose to Collaborate?**Kaiwen Li, University of Southern California, Los Angeles, CA, United States, Mingxi Zhu**

This paper studies how the interplay between market power and data volume shapes firms' incentives to collaborate via federated learning (FL) platforms. While prior research often models information sharing as stylized signaling, it overlooks the complexity and scale of real-world customer data. Federated learning offers a practical alternative, allowing firms to jointly learn key market parameters without revealing raw data. We develop a game-theoretic model in which competing firms strategically decide whether to participate in FL. Our analysis identifies precise conditions under which data-rich or data-poor firms—depending on their relative market positions—choose to collaborate. We show that collaboration is more likely when data asymmetries outweigh market asymmetries, revealing a fundamental trade-off between data power and market power. These findings offer design principles for collaborative learning platforms in competitive markets and provide insights into when cooperation emerges among rivals.

3 - Building Agents for Healthcare Workflow Automation**Junze Ye, Stanford GSB, Stanford, CA, United States, Jiayi Wang, Mohsen Bayati**

We present an agentic framework for the automation of clerical and diagnostic tasks in healthcare, and a data-efficient way of finetuning an LLM to improve its performance on such task. An MDP model is presented to capture the sequential decision-making nature of such a workflow, and we validate the effectiveness of our approach on both real-world and synthetic data.

4 - AI and Machine Learning at Scale in Etsy's Marketplace**Salomón Wollenstein Beteck, Etsy, Washington DC, DC, United States**

How does a global marketplace for unique and creative goods harness the power of artificial intelligence? In this session, we explore how Etsy leverages cutting-edge machine learning and data science to enhance search relevance, recommendation systems, pricing strategy, and fraud detection—while preserving the human touch at the heart of its community. Featuring an industry expert from Etsy, this talk offers a behind-the-scenes look at the real-world challenges and solutions in deploying AI at scale in a two-sided marketplace. Attendees will gain insights into how Etsy balances personalization, fairness, and business objectives using applied ML, and what it takes to translate research into production in a dynamic e-commerce environment.

MD55

Building B Level 3 B314

Pricing, Learning and Optimization in Scale

Invited Session

Revenue Management

Chair: Mengzhenyu Zhang, University College London, London, United Kingdom

Co-Chair: Yi-Chun Akchen, University College London, London, United Kingdom

1 - Beyond $O(\sqrt{T})$ Regret: Decoupling Learning and Decision-Making in Online Linear Programming**Wenzhi Gao, Stanford University, Stanford, CA, United States, Dongdong Ge, Chunlin Sun, Chenyu Xue, Yinyu Ye**

Online linear programming plays an important role in both revenue management and resource allocation, and recent research has focused on developing efficient first-order online learning algorithms. Despite the empirical success of first-order methods, they typically achieve a regret no better than $O(\sqrt{T})$, which is suboptimal compared to the $O(\log T)$ bound guaranteed by the state-of-the-art linear programming (LP)-based online algorithms. This paper establishes a general framework that improves upon the $O(\sqrt{T})$ result when the LP dual problem exhibits certain error bound conditions. For the first time, we show that first-order learning algorithms achieve $o(\sqrt{T})$ regret in the continuous support setting and $O(\log T)$ regret in the finite support setting beyond the non-degeneracy assumption. Our results significantly improve the state-of-the-art regret results and provide new insights for sequential decision-making.

2 - On the Role of Curricula in the Decentralization of Service Delivery at Scale**Christopher Ryan, University of British Columbia, Vancouver, BC, Canada**

We discuss several case studies of service organizations using decentralization – mediated by curriculum development – in their process to reach scale. We will draw conclusions about the important operational role curriculum development can play in scaling service operations and offer an outline for future potential research.

3 - Dynamic Pricing for Last-Mile AI Automation: a Data-Driven Approach with Temporally Shrinking Choice Sets**Jingwen Tang, University of Miami, Coral Gables, FL, United States, Shuai Hao, Yuqian Xu, Cong Shi**

Motivated by one company offering autonomous on-demand delivery service, we study the dynamic pricing problem for the delivery time windows with capacity constraints. We formulate the problem into a multiproduct pricing problem, with a shrinking choice set, and the consumer choice is captured by an MNL model. To overcome the challenge raised by the diminishing delivery time windows, we decompose the system into multiple subsystems, which are linked with the capacity constraints. We propose a heuristic pricing strategy and show its asymptotic optimality in revenue as the problem scale (customer arrival rates, capacities) increases. We show the proposed strategy increases the revenue significantly compared to the current pricing policy by numerical experiments using the real data of the company. The method not only provides the practitioners with a practical pricing strategy with asymptotic optimal performance but also lends insights into other similar problems with varying choice sets.

4 - Evaluating Scheduling Policies with Lower Bias: An Arrival-Based Approach for Switchback Designs**Jierui Zuo, Tsinghua University, Beijing, China, People's Republic of, Jinglong Zhao, Zijie Zhou**

We propose an arrival-based estimator for switchback experiments designed to compare scheduling policies—specifically First-Come-First-Serve (FCFS) and Shortest-Job-First (SJF). Under a stationary workload assumption, where job arrival and job length distributions remain fixed, the commonly used naive estimator exhibits systematic bias arising from interactions among job arrival patterns, scheduling decisions, and queue dynamics. Our arrival-based estimator explicitly incorporates job arrivals to correct for this bias. We provide theoretical analysis demonstrating that our estimator achieves strictly lower expected bias compared to the naive estimator.

Simulation results further confirm that the arrival-based estimator yields more accurate and reliable comparisons of scheduling policies. Our method is applicable in many practical applications, such as LLM inference, where precise policy evaluation is critical for performance optimization.

MD56

Building B Level 3 B315

Dynamic Matching and Resource Allocation

Invited Session

Auctions and Market Design

Chair: Süleyman Kerimov, Rice University, Houston, TX, United States

Co-Chair: Sophie Yu, The Wharton School of the University of Pennsylvania, Philadelphia, PA, United States

1 - Courier Ranking and Pricing

Matias Romero, Columbia Business School, New York, NY, United States, Hongyao Ma, Will Ma

On-demand delivery platforms typically fulfill orders by sending sequential take-it-or-leave-it offers to couriers, specifying both the job and the payment. This process involves ranking couriers and setting payments along the sequence of offers. In practice, these decisions are often made by different teams, each focusing on a single lever, which raises the question of how ranking and pricing interact. We study a parsimonious model that captures this interaction, where the platform knows both the value of matching each courier to a job and the acceptance probability induced by any payment. Our findings reveal that independently optimized decisions can lead to inefficiencies, and we show that simple policies achieve global optimality in important cases.

2 - Online Selection with Uncertain Disruption

Yihua Xu, Rice University, Houston, TX, United States, Süleyman Kerimov, Sebastian Perez-Salazar

In numerous online selection problems, decision-makers (DMs) must allocate on the fly limited resources to customers with uncertain values. The DM faces the tension between allocating resources to currently observed values and saving them for potentially better, unobserved values in the future. Addressing this tension becomes more demanding if an uncertain disruption occurs while serving customers. Without any disruption, the DM gets access to the capacity information to serve customers throughout the time horizon. However, with uncertain disruption, the DM must act more cautiously due to risk of running out of capacity abruptly or misusing the resources. Motivated by this tension, we introduce the Online Selection with Uncertain Disruption (OS-UD) problem. In OS-UD, a DM sequentially observes n non-negative values drawn from a common distribution and must commit to select or reject each value in real time, without revisiting past values. The disruption is modeled as a Bernoulli random variable with probability p each time DM selects a value. We aim to design an online algorithm that maximizes the expected sum of selected values before a disruption occurs, if any.

We evaluate online algorithms using the competitive ratio. Using a quantile-based approach, we devise a non-adaptive single-threshold algorithm that attains a competitive ratio of at least $1-1/e$, and an adaptive threshold algorithm characterized by a sequence of non-increasing thresholds that attains an asymptotic competitive ratio of at least 0.745. Both of these results are worst-case optimal within their corresponding class of algorithms.

3 - Optimality of Random Regular Graphs in Sparse Operations Designs

Xiaochun Niu, Duke University, Durham, NC, United States, Weijia Li, Yehua Wei, Jiaming Xu

The problems of designing sparse graphs arises frequently in resource allocation and operations research. For example, designing sparse flexible processes is widely used to effectively handle uncertain demand in production systems. Here the goal is to construct the sparsest bipartite graph between supply and demand sides, while achieving at least $(1-\epsilon)$ of the maximum sales of a fully flexible system. In addition, in middle-mile delivery operations, designing sparse subgraphs that sustain large matchings after random node deletions helps maintain predictable delivery routes. In this context, the goal is to design the sparsest graph whose maximum matching size is at least $(1-\epsilon)$ times that of the fully connected graph.

There has been a growing literature studying graph designs for both problems, where the performance of K -clusters, K -chains, and Erdős-Rényi graphs have been extensively analyzed. However, optimal sparse graph designs have remained an open question. In this paper, we prove that random d -regular graphs achieve $(1-\epsilon)$ -optimality for both problems. Specifically, its node degree d matches the lower bound on the average node degree required for any graph to achieve $(1-\epsilon)$ -optimality, thereby establishing the optimality of random d -regular graphs in these sparse design problems.

We also provide optimal bounds on the average node degree for graph designs involving geometric structure, such as K -chains and one-dimensional random geometric graphs, which characterize the location or feature vectors of nodes. Numerical experiments are included to enhance our theoretical findings.

4 - Matching Queues with Strategic Agents

Hossein Dabirian, University of Michigan, Ann Arbor, MI, United States, Saif Benjaafar, Brian Denton, Mustafa Akan, Philipp Afèche

We consider a two-sided queue where agents (e.g., customers and service providers) arrive continuously over time to each side, seeking to be matched with each other. In deciding whether to join the queue, agents are sensitive to the reward they earn upon matching and the cost incurred while waiting. It is well known that self-optimizing behavior does not maximize overall welfare in one-sided queues [Naor, 1969]. We extend this result to two-sided queues and propose a tolling strategy to align individual incentives with the welfare-maximizing strategy. In particular, we show that this may involve imposing a fee on one side (charging a positive price) while offering a subsidy for the other (a negative price). In contrast to results for one-sided queues, we show that the socially-optimal limits on queue sizes may be smaller or larger than those under self-optimization. We also characterize the price of anarchy, providing a tight bound.

5 - Managing Reusable Resources with Continuous Arrivals and Discrete Usage

Yun-Tung Kuo, Duke University, Durham, NC, United States, Cong Shi, Yehua Wei

In the hotel and car rental industries, resources are reusable, and the usages of these resources are often reserved in advance. While extensive research exists on managing reusable resources, a critical yet often overlooked feature in these industries is their dual time-scale dynamic. Specifically, reservation arrivals occur continuously over time, whereas usage times, characterized by fixed check-in and check-out times, are discrete.

We propose a dual time-scale dynamic model to seek to understand how reusable resources can be efficiently managed in the presence of advanced reservation arrivals. Surprisingly, while the dual time-scale dynamic disrupts the stationarity of usage requests, it enables the model to be closely approximated by offline resource allocation problems. This insight allows us to develop a state-dependent admission policy that achieves an optimality scaling superior to existing approaches that do not account for the dual time-scale feature. Our findings demonstrate that embracing the dual time-scales not only overcomes challenges but also strengthens theoretical guarantees, thus unlocking new strategies for managing reusable resources.

MD57

Building B Level 3 B316

Analytics in Social Media and Digital Technology

Invited Session

Social Media Analytics

Chair: Zefeng Bai, 24450

1 - Change Points in Functional Time Series Using Combinations of Strong Harmonics: Applications to Measuring Social Media Impacts of Sponsoring Large Sporting Tournaments

Moinak Bhaduri, Bentley University, Waltham, MA, United States

Discrete time series, owing to interrupted sampling, do not make provisions for a variable's existence in the intervening space between adjacent time points. The problem becomes especially acute for high-frequency data, such as (positive, negative, neutral) tweets of companies that lend their sponsorship to large sporting events like the US Open. In this talk, we examine how non-functional versions of common statistical objects like principal components upgrade to their functional counterparts and help us with the analogous tasks in functional settings: detecting change-points, for instance. We deploy inner products of crucial harmonics to query full functions (instead of parameters) and test whether between companies that do and do not sponsor the large events such as the US Open, break points occur at similar times around the tournament, and are of similar intensities. Thanks to our PhD student Shardul Wartikar for collecting sentiment data on tweets for sixteen sponsoring and sixteen non-sponsoring companies.

2 - Designing a Persuasion-Aware Smishing Detector for Evolving Threats

Chen Zhong, The University of Tampa, Tampa, FL, United States, Wei Xie

Smishing, or SMS-based phishing, continues to evolve as attackers increasingly exploit large language models (LLMs) to generate varied and convincing messages that alter tone, language, and keywords. This adaptability creates a significant gap: the lack of effective, explainable detection systems that can adapt to different smishing styles and strategies. Addressing this gap, we adopt a design science method to design and develop an effective smishing detection system grounded in persuasion theory. Building on our prior work, where we manually analyzed and coded 934 smishing messages according to persuasive techniques, this study leverages that dataset to train a machine learning model capable of identifying these strategies. Specifically, we implemented a RoBERTa-based classifier to detect persuasion cues with high accuracy. To promote transparency and foster user trust, we integrated Explainable AI (XAI) techniques that highlight which textual features contribute to the model's decisions. This feature provides users with meaningful explanations that can support long-term learning and awareness. Furthermore, we explore the tradeoff between accuracy and explainability, emphasizing the importance of a balanced approach that maintains detection performance while offering interpretable results. Our work contributes a replicable method for smishing detection grounded in theory and user-centered design, offering practical implications for both researchers and cybersecurity practitioners.

3 - Influenced by Influence: the Role of Social Media In Financial Knowledge Gaps

Jinglin Hu, University of Florida, Gainesville, FL, United States, Zefeng Bai, Ying Wang, Zimou Xiang

As more people rely on social media for financial advice, concerns have grown over its impact on individuals' understanding of personal finance. This study examines the relationship between social media use and financial knowledge disparity-- the gap between subjective and objective financial knowledge. Using data from the 2018 and 2021 waves of the National Financial Capability Study (NFCS), we find that individuals who seek investment advice on social media exhibit significantly greater financial knowledge disparity. Interestingly, the disparity is even more pronounced among individuals with a background in financial education. Our follow-up analysis confirms that greater financial knowledge disparity is associated with a higher likelihood of undesirable financial behaviors, including credit card misuse, early retirement withdrawals, and delayed mortgage payments.

These findings underscore the need for more effective, targeted financial education and stronger oversight of online financial content. By identifying vulnerable populations and the risks associated with financial knowledge disparity, our study can help policymakers, educators, and social platforms develop more informed interventions to encourage responsible financial behavior in this digital era.

4 - Corporate DEI Messaging on Social Media: Effects of Authenticity and Political Ideology on Consumer Engagement

Lingshu Hu, Washington and Lee University, Lexington, VA, United States, Weilu Zhang, Rui Sun, Detelina Marinova, Nathanael Seife

In recent years, brands and companies have renewed their commitments to diversity, equity, and inclusion (DEI) in response to evolving societal expectations. However, DEI-related content can be controversial and highly polarized, meaning its effects on consumer digital engagement may vary depending on the context. Drawing on Twitter data from approximately 500 U.S.-based firms, this study investigates the impact of engaging with DEI-related issues as part of corporate social media strategy. Specifically, we examine how political ideology and message authenticity influence the effectiveness of DEI-related tweets.

MD58

Building B Level 3 B317

Optimization for Scheduling and Project Management

Contributed Session

Contributed

Chair: Julien Maurice, Georgia Institute of Technology, Physical Internet Center, Atlanta, GA, United States

1 - Fulfillment Scheduling with a Concave Shipping Cost Structure

Liangzi Zhu, University of Maryland, College Park, MD, United States, Zhi-Long Chen

We study a fulfillment scheduling problem under a nondecreasing concave shipping cost structure. Unlike traditional linear cost models, our shipping cost function more closely reflects the real-world delivery cost. Each order must be processed first and then delivered to its respective customer by a committed deadline. An unlimited number of homogeneous vehicles is available for delivery, each with a capacity for a fixed number of orders. Our goal is to find a fulfillment schedule that minimizes the total delivery cost while ensuring that all order deadlines are met.

We analyze different cases in this study. With one customer, one processing line, and orders having agreeable processing times and deadlines, the problem can be solved to optimality by a dynamic programming algorithm in polynomial time. In addition, with multiple customers, identical parallel processing lines, and orders having non-agreeable processing times and deadlines, this problem is strongly NP-hard. We tackle this case using a column generation based heuristic to obtain near-optimal solutions.

2 - Robust Optimization in Generalized Project Networks

Siyuan Chen, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Nicholas Hall, Miao Song

The growing prevalence of project management as a business process is matched by its growing range of applications. Notably, modern projects are characterized not only by customary task time uncertainty, but also by uncertainties about their scope and constraints than traditional ones. However, research has so far not incorporated several of these fundamental uncertainties into project optimization. We address this deficiency by incorporating the modern project realities of (a) uncertain full or partial task rework, (b) the unexpected arrival of new tasks, (c) uncertainty in the existence of task precedence relationships, (d) probabilistic branching between alternative plans, and (e) uncertain availability or feasibility of contingency plans, into a general project network model. This model minimizes the total of plan evaluation and plan execution costs as well as expected tardiness relative to a defined makespan target. The model is solved by backward dynamic programming, using the sequential evaluation of contingency plans to define stages. The evaluation of tardiness at the terminal stage is accomplished by a distributionally robust optimization model. A computational study shows that the overall model is solvable within reasonable running time for up to 150 project tasks. We also use sensitivity analysis to identify insights for projects

managers. Our work expands the range of project management applications for which optimization is realistically possible. A special case where each plan has no more than three contingencies is shown to be more efficiently solvable.

Key words: project management, uncertain scope and constraints, dynamic programming, distributionally robust optimization.

3 - Poisoned Apples: Project Team Experiential Diversity, Project Performance, and Voluntary Turnover in the Context of It and Consulting Services

Yunfan Wang, IE Business School, Madrid, Spain, Fabrizio Salvador, Emmanouil Avgerinos, Marco Caserta

Retaining skilled human capital is essential for project performance in knowledge-intensive businesses. To this end, firms can assign employees to experientially diverse project teams, providing them with learning opportunities that foster motivation and commitment to the firm. However, as learning increases employees' market value, it can paradoxically raise the risk of voluntary turnover. This conundrum is exacerbated when considering the effect of team experiential diversity on project performance: promoting diversity may create a more engaging work environment but is incompatible with having stable teams of specialists. In this study, we investigate the effect of team experiential diversity using data from 810 project teams, involving 3,139 employees, from a multinational consulting and IT services firm. We focus on team experiential diversity relative to customer experience (i.e., client experience diversity) and familiarity among team members (i.e., coworker experience diversity). We find that team experiential diversity increases project performance but has an inverted U-shaped relationship with voluntary turnover. Furthermore, we use the variables from our regression models to build predictive machine learning models for project performance and voluntary turnover. We find that after including the team experiential diversity variables, model accuracy increases for both models, further supporting their economic importance. With this paper, we contribute to broadening the focus of operations management research on project teams beyond project performance to include talent retention—a fundamental goal and concern for project managers in knowledge-intensive industries.

4 - Cost Optimization of Repetitive Project Scheduling with Integrated Material Ordering

Xiaoqi Lei, Huazhong University of Science and Technology, Wuhan, China, People's Republic of, Nanfang Cui

With the rapid growth of prefabricated construction, efficient material management and project scheduling have become critical challenges. Key materials like prefabricated components, concrete, and steel reinforcement involve substantial transportation and storage costs, significantly impacting overall project expenses. Since project scheduling and material ordering are interdependent, integrated optimization is crucial for cost efficiency. However, existing research on the Project Scheduling and Material Ordering Problem (PSMOP) focuses on general projects, neglecting repetitive projects which involve complex network relationships, work continuity constraints, etc.

This study analyzed the trade-offs between material ordering costs, inventory costs, and interval time costs in the **Repetitive Project Scheduling and Material Ordering Problem (RPSMOP)**. A **0-1 integer linear programming model** is formulated, optimizing activity start times and material ordering schedules to minimize overall costs. To solve the model, a **nested simulated annealing genetic algorithm (SAGA)** based on the encoding of activity start time was designed. Its core is to determine the activity time window and generate feasible solutions using different rules to handle the continuity constraints. The validity and superiority of the model and algorithm were verified through large-scale simulation and sensitivity analysis. The solution results for four different hypothetical scenarios show that under the condition of interruptible work, adjusting the start time of non-critical activities can obtain better project scheduling and material ordering plans. The sensitivity analysis of cost parameters shows that the algorithm optimization is mainly achieved by merging orders to reduce the number of orders.

5 - A Heuristic Approach to Scheduling Raw-Material Unloading at A Steelworks

Hansol Kim, Pohang University of Science and Technology (POSTECH), Pohang, Korea, Republic of, Seonggyun Hong, Byung-In Kim, Hae-Jin Sung, Hyeonon Wi

The raw-material unloading scheduling problem involves determining an efficient schedule for unloading iron ore and coal from ships at a steelworks equipped with six discrete berths. The complexity of this problem is heightened by several factors, including varying unloader capacities based on both unloader type and carried raw material brands, as well as eligibility constraints between berths and raw materials. Additionally, since the unloaded materials must be transported to designated storage yards via belt conveyor systems, where no two routes can simultaneously use a common conveyor link, the complexity of conveyor networks and transportation planning must also be considered. If a ship's unloading is completed later than its designated time, a demurrage cost is incurred. The objective of this scheduling problem is to minimize the total demurrage cost. However, due to uncertainties such as weather and wind speed, the unloading time of a ship cannot be precisely estimated solely from the material types, quantities, and unloader specifications. To address this, we propose a linear programming model to estimate ship unloading times. We also developed an iterative heuristic algorithm to efficiently solve the unloading scheduling problem. The proposed algorithm is evaluated using real-world data from the world's largest steelworks, demonstrating both its effectiveness and practical applicability. The algorithm is currently undergoing pilot testing at the steelworks.

6 - Integrated Optimization of Steel Grade Assortment, Production, and Inventory Planning

Gongshu Wang, Northeastern University, Shenyang, China, People's Republic of, Sibao Liu, Lixin Tang

In response to increasingly diversified and personalized market demands, the steel industry is shifting from traditional mass production to mass customization in order to enhance production flexibility, reduce costs, and better meet individual customer requirements. This study addresses the integrated optimization of steel grade assortment, production planning, and inventory control within a two-stage manufacturing system specifically configured to enable mass customization. The problem entails selecting appropriate steel grades for customer orders, determining production quantities for both intermediates and finished products, and managing intermediate inventories—aiming to minimize total costs while maximizing order fulfillment within delivery windows. The complexity arises from a two-level decision-making structure, where high-level decisions on steel grade assortment are interdependent with low-level production and inventory planning, creating intricate coordination challenges across stages. To address this, we formulate a mixed-integer programming model and develop a tailored metaheuristic approach. The method combines: (i) hierarchical optimization for constructing initial solutions, (ii) a fix-and-optimize strategy for re-optimization of decomposed variable blocks, and (iii) a kernel-based intensification mechanism that iteratively refines promising regions of the solution space. Computational experiments on synthetic instances demonstrate the effectiveness of the proposed approach and offer managerial insights into achieving a balance between cost efficiency and order responsiveness in customized steel production.

翻译

搜索

复制

7 - Toward a Resilient Project Configuration and Planning: A Proactive-Reactive Optimization Approach Under Complex Disruptions

Julien Maurice, Georgia Institute of Technology, Atlanta, GA, United States, Jihye Jung, Benoit Montreuil, Simon Kwon, mahsa mahdavisarif

While traditional project planning focuses on assigning start times to activities, it often relies on simplified, independent activity configurations (small set of predefined options that only affect individual durations or costs). In contrast, project configuration extends this scope by accounting for interdependencies between activity configurations throughout the project lifecycle (e.g. production rate, storage and transportation strategy, or resource level). This approach is particularly valuable in complex projects, such as in the construction industry, where activities are highly interdependent, subject to disruptions and involve multiple stakeholders requiring dynamic decision support.

However, most existing methods in the literature simplify these dependencies and consider only basic disruptions scenarios. To address this gap, we propose a Proactive-Reactive Project Configurator and Planning that supports both baseline planning and adaptive reconfiguration.

Our framework first builds a baseline configuration using a mixed-integer programming (MIP) model that determines each activity's execution mode and start time, with the objective of minimizing total project cost and completion time. Then, different levels of disruptions under various scenarios are simulated to reflect real-world uncertainties. These scenarios are clustered based on their impact on model outputs and decision variables for different stakeholders. This method enables the identification of recurring reconfiguration strategies and supports robust, resilient reconfiguration of the project time and cost.

We demonstrate the applicability of our approach on a complex construction project with a modular and distributed structure. We also analyze the resiliency of the project configuration on adapting to disruptions and guiding decision-making across both proactive and reactive phases.

MD59

Building B Level 3 B318

Healthcare Systems

Contributed Session

Contributed

Chair: Maryam Zakeri, Alberta School of Business, Edmonton, AB, Canada

1 - Development and Cost Effectiveness of a Referral Process for Integrated Medical and Dental Care

Tasfia Mubashshirath Roshnee, The Pennsylvania State University, University Park, PA, United States, Paul Griffin

In the US, there has been a lack of integration between medical and dental care for chronic conditions such as Type 2 Diabetes Mellitus (T2DM). There is an opportunity to improve patient outcomes through better coordination. For example, tooth scaling and root planning for patients with periodontal disease has been shown to significantly reduce HbA1c levels. We develop a patient referral model and consider its cost effectiveness for patients with T2DM. We first use a Bayesian network (BN) model to predict disease progression and complication risks among T2DM patients utilizing longitudinal patient data from electronic health records (EHR) to identify high-risk T2DM patients who would benefit from timely dental screening and intervention through better glycemic control and reduced likelihood of developing a diabetic-related complication such as neuropathy or diabetic retinopathy. We consider clinical, demographic, and behavioral variables to provide patient-specific risk assessments for periodontal disease. We then consider the cost-effectiveness of the proposed referral strategy from the payer perspective. We conduct sensitivity analyses and compare the results with actual observed clinical outcomes. The overall goal is to develop a framework to improve clinical outcomes and optimize resource allocation by effectively preventing and mitigating diabetes-related complications.

2 - A Cross-Continental Investigation of Race-Based Variations in Leukemia Drug Toxicity Using OpenFDA Pharmacovigilance Data

Hossein Amini, Washington University in St. Louis, Saint Louis, MO, United States, Ali Abbasi

This study investigates the diversity of responses to FDA-approved leukemia treatments by analyzing data from the OpenFDA adverse event report system. We examine a comprehensive dataset comprising over 300,000 safety reports submitted between 2004 and 2022, allowing us to assess trends over time and across various global regions. Utilizing a range of analytical methods, we identify patterns in adverse reactions and explore whether the timing of side effects differs among various racial groups.

Our research findings indicate that certain populations may experience more pronounced or earlier onset of side effects from leukemia medications. For example, specific racial groups exhibit higher rates of particular adverse reactions in designated regions, and we have observed variations in the timing of side effect onset among other demographics. These patterns emphasize the necessity of considering racial and geographic factors when evaluating drug safety and efficacy. Ultimately, our study underscores the imperative for more personalized approaches in the monitoring and management of leukemia treatment to promote improved and more equitable health outcomes globally.

3 - Temporally Feathered Radiation Therapy Planning

Aysenur Karagoz, Rice University, Houston, TX, United States, M. Soheil Hemmati, Andrew Schaefer

Intensity-modulated radiation therapy (IMRT) remains the standard-of-care radiation therapy technique for most head-and-neck cancer (HNC) patients. Despite the relatively high survival rate of HNC (65-75%), the survivors often suffer from post-treatment long-lasting side effects, which negatively affect their quality of life. In an IMRT-based treatment plan, the daily doses delivered to the tumor and organs-at-risk (OARs) remain fixed based on the premise that the one-day inter-fraction suffices for each OAR's recovery. However, a novel treatment planning approach, referred to as temporally feathered radiation therapy (TFRT), proposes variable daily dose delivery to the OARs, aiming at increasing the recovery time of OARs at the cost of receiving higher doses once a week. While initial implementations of TFRT have shown promising results in the clinic, they do not distinguish between the inherent radiobiological differences among OARs for HNC patients, thus prescribing longer, yet uniform recovery time for all OARs. By considering variable radiosensitivity profiles across multiple OARs for HNC radiation therapy, we develop a mixed-integer nonlinear program (MINLP) that computes specific recovery time for each OAR to minimize the toxicity burden to the patients. Our proposed MINLP leverages the linear-quadratic model, which predicts each OAR's survival from radiation-induced damages. Our model is further flexible in addressing specific prioritization of OARs based on the clinician's preferences. The result of our model indicates a lower overall toxicity burden when OARs with higher radiosensitivity have longer recovery times under the TFRT framework.

4 - Physician Workload and Medical Malpractice Liability: An Empirical Investigation

Denise Bergdolt, University of Pittsburgh, Pittsburgh, PA, United States, Prakash Mirchandani

Medical malpractice—professional negligence by healthcare practitioners resulting in patient harm—is a critical issue in the U.S. healthcare system, affecting thousands of stakeholders and carrying a financial impact of billions of dollars annually. We construct an extensive, national, retrospective database to examine how physician workload intensity impacts medical malpractice case intensity at a state-year level.

To investigate the relationship, we implement a generalized propensity score stratification method suited for continuous workload intensities. We find a concave relationship between the natural log of the workload intensity and case intensity. At low workload intensity levels, increases in workload intensity are associated with rising medical malpractice case intensity levels. Counterintuitively, beyond a tipping point, as workload intensity further increases, case intensity decreases.

We attribute the concave pattern to a shift in the relative dominance of the work stress and defensive medicine effects. As workload rises, physicians may experience increased stress, mental workload and burnout, all of which negatively correlate with medical setting job performance in the literature. Meanwhile, a different stream of literature illustrates that if physicians do not have sufficient time to inspect their patients, they are more likely to use defensive medicine, which leads to increased treatment documentation, ultimately decreasing the odds of medical malpractice conviction in court settings.

We conduct robustness checks using alternative model specifications, and sensitivity analysis to account for unobserved covariates. We discuss the practical and theoretical impact of our findings on medical malpractice mitigation, and propose future streams of research.

5 - Adaptive Behavior of Paramedics: Impact of Paramedic Crew Workload on Scene TIME Duration and Probability of Patient Transport

Maryam Zakeri, Alberta School of Business, Edmonton, AB, Canada, Armann Ingolfsson, Mohammad Delasay, Kenneth Schultz, Graham Vanderwater, Gerald Lazarenko, Jim Garland, Peter Helfrich

Emergency medical services (EMS) play a critical role in patient care, yet the variability in demand—by severity, timing, and location—makes capacity planning challenging. Paramedics must stabilize, treat, refer or transport patients while managing both physical and cognitive fatigue, which may influence their decision-making. This research explores how fatigue affects EMS crew decisions on scene. We develop a workload measure that represents of fatigue and employ statistical methods to establish a causal link between workload and scene outcomes. Specifically, we analyze scene time and transport decision across four call types: cardiac arrest, trauma, opioid overdose, and influenza-like illness. Our findings provide insights for optimizing EMS resource management and improving patient care efficiency.

MD60

Building B Level 3 B319

Human and Societal Applications of AI

Contributed Session

Contributed

Chair: Guoxin Wang, Renmin University of China, Beijing, N/A

1 - Challenges and Safety of AI Applications

Young Lee, UL Solutions, Melville, NY, United States, Sudhi Sinha, Kenneth Boyce

Artificial Intelligence (AI) is driving transformative changes across industries by enhancing automation, predictive analytics, intelligent decision-making, and operational efficiency. However, the journey from AI development to real-world deployment is riddled with significant challenges. This presentation delves into two critical facets of AI adoption in the industrial sector: the multifaceted challenges that organizations encounter while implementing AI solutions and the paramount importance of AI system safety to ensure reliable, ethical, and secure functionality.

The talk covers the complexities of industrial AI development, including data collection and management, model selection, deployment scalability and system integration. These challenges not only impact the performance and success of AI applications but also raise concerns regarding trust, transparency, and accountability.

Furthermore, the presentation will address the risks linked to AI model robustness and reliability, the need for transparency and explainability, control and oversight mechanisms, biased decision-making, data privacy concerns, security vulnerabilities, regulatory compliance, and broader societal impacts. The talk examines how safety considerations are integral to addressing deployment challenges, forming a bridge between AI reliability, ethical responsibility, and regulatory adherence.

By exploring these interconnected themes, the talk aims to provide insights into the evolving landscape of AI adoption, highlighting the need for a balanced approach that embraces both innovation and responsibility in shaping the future of AI-powered industries.

2 - Real-Time Bilevel Contingency Scheduling via AI-Inspired Adversarial Search

Adam DeHollander, University at Buffalo, Buffalo, NY, United States, Moises Sudit, Mark Karwan, John Crassidis

We study a bilevel scheduling problem in which a “leader” must complete a set of interdependent tasks while an “adversary” can enact disruptive moves that delay task execution. This formulation captures a broad class of real-world contingency planning and resilience challenges. Solving the resulting bilevel optimization exactly is intractable in most instances. We adapt adversarial search techniques from board games such as chess (e.g., lower bounds, pruning, and transposition tables) to dramatically reduce the decision space. On benchmark instances, our AI-inspired solver demonstrated orders of magnitude improvement in the time to solve the problem and achieve near-optimal solutions by decreasing the decision space, making it practical for rapid, real-time planning.

3 - How Generative AI Reviews Shape Online Engagement

Wenyu Jiao, University of Science and Technology of China, Hefei, China, People's Republic of

Recent advancements in generative AI for content creation present significant opportunities to transform the online retail industry. This study examines the effects of generative AI-enabled reviews on customer experience and engagement. In collaboration with an e-commerce platform and using a quasi-experimental design, we find that generative AI-generated reviews enhance customer behavior by increasing both viewing time and purchase rates. However, they also lead to a reduction in customer engagement in the form of review contributions. Our analysis reveals that these AI-generated reviews are particularly effective in boosting sales within functional product categories and have a greater positive impact when a diverse set of reviews is available. Additionally, they significantly improve the experience of less experienced online shoppers. These findings offer valuable insights for e-commerce platforms implementing AI-powered content creation tools.

4 - Extracting Policy Intelligence from Public Data Sources : LLMs as Policy Assistants

Giyun Kim, Seoul National University, Seoul, Korea, Republic of, Sanghyun Park, Dongyoon Shin, Jungeun Kim, Sungjoo Lee

Developing domain-specific frameworks for large language model (LLM) applications has been widely studied to enhance research efficiency across various fields. As policy planning requires detailed analysis of reliable sources, LLM applications have significant potential to enhance analytical efficiency. Therefore, we propose a novel LLM application framework designed for policy planning that reduces the time and resources required for analyzing global policy trends and their comprehensive content, including goals, implementation processes, outcomes, and causal relationships. First, a systematic collection and evaluation of datasets to identify reliable sources of global policy information that could serve as valuable references for policymakers was conducted. This process involved indicator development for data quality assessment in terms of reliability, relevance, completeness, and timeliness. We then categorized types of policy planning based on an in-depth analysis of previous policy documents, establishing key content requirements for each category. Subsequently, the fine-tuning and retrieval-augmented generation (RAG) approaches for the datasets were suggested and validated to efficiently extract required information from the collected datasets. We expect the framework to help policymakers effectively analyze global policy landscapes, identify policy implementation strategies along with their results, and derive valuable insights for policy planning across diverse contexts and challenges.

5 - Fast vs. Wise: Effect of Decision Speed and Quality on Users' Adoption of AI Advice

Guoxin Wang, Renmin University of China, Beijing, China, People's Republic of

Drawing on cue utilization theory, this research investigates how decision speed and quality affect users' adoption of artificial intelligence (AI) advice. Three experiments demonstrate that users exhibit higher adoption intention when AI decision speed or quality is high versus low, mediated by perceived intelligence and perceived risk. When decision quality is high, the positive impact of decision speed is attenuated. Importantly, task complexity and time pressure jointly moderate these relationships. Under high time pressure, users prioritize decision speed over quality regardless of task complexity, as speed becomes a dominant cue for AI competence. However, under low time pressure, task complexity determines cue utilization: for simple tasks, speed remains influential, while for complex tasks, quality becomes the primary determinant of adoption intention. The research extends cue utilization theory to human-AI interaction, demonstrating how situational factors shape the relative importance of AI performance cues. Practical implications include designing adaptive AI interfaces that emphasize speed indicators during time-pressured situations and quality indicators for complex tasks in reflective contexts, optimizing user adoption across diverse decision environments.

MD61

Building B Level 4 B401

Uncertainty-Aware Surrogate Modeling and Optimization for Complex Systems

Invited Session

Quality, Statistics and Reliability

Chair: Kangan Chen, University of Wisconsin Madison, Madison, WI, United States

Co-Chair: Siqin Yang, Arizona State University, Tempe, AZ, United States

Co-Chair: Andi Wang, University of Wisconsin-Madison, Madison, WI, United States

1 - Exact Multistage Bayesian Optimization**Kangan Chen, University of Wisconsin-Madison, Madison, WI, United States, Siqin Yang, Hao Yan, Andi Wang**

This paper introduces the Exact Multistage Bayesian Optimization (EMSBO) algorithm for optimizing multistage experiment, where intermediate outputs serve as part of subsequent stage inputs. Unlike existing approaches that either ignore intermediate observations or rely on heuristic methods, EMSBO employs a dynamic programming formulation to exactly compute the expected utility of each stage considering all subsequent stages. The algorithm's utility functions are approximated through posterior Monte Carlo sampling and nearest neighbor interpolation. The theoretical bounds are established for the approximation error. Experimental results demonstrate that EMSBO outperforms existing methods like BOFN and CASCADE, particularly in early optimization iterations, while it maintains computational scalability for problems with up to 10 stages. The method's effectiveness stems from exploiting the sparse dependency structure inherent in multistage experiments, enabling exact solutions where broader look-ahead Bayesian optimization approaches become intractable. This makes EMSBO particularly valuable for optimizing complex sequential processes in scientific and engineering applications.

2 - Functional Approximated Multistage Bo: Reformation of Multistage Bo in RL Fashion**Siqin Yang, Arizona State University, Mesa, AZ, United States, Hao Yan, Zhenyu Kong**

Multistage experiments are commonly observed in industrial applications. While current composite Bayesian Optimization methods utilize intermediate results to provide better performance than standard BO, we propose a principled framework that reformulates multistage Bayesian Optimization as a Markov Decision Process. Under this formulation, we employ dynamic programming techniques with approximations to address the problem. Our Functional Approximated Multistage Bayesian Optimization (FAMBO) algorithm efficiently makes advantage of functional approximation and on-policy sampling. Empirical experiments show that our algorithm outperforms or at least performs considerably well compared to SOTA methods with much better performance in early stages and shorter computational time.

3 - Temporal Causal Representation Learning for Distributed Data**Jianhong Chen, northeastern university, Everett, MA, United States**

Causal representation learning, also known as causal discovery, has garnered attention for analyzing observational data because controlled experiments can be costly. In this setting, Bayesian Networks (BNs) are the standard score-based approach for learning causal diagrams. However, BNs capture only contemporaneous causal relations. To model time-lagged relations, researchers have extended BNs into Dynamic Bayesian Networks (DBNs). Traditionally, structure learning for Dynamic Bayesian Networks has been performed in a centralized fashion, requiring all data to be pooled in one location. However, in real-world settings, data are often distributed across multiple entities (e.g., companies) that wish to collaborate on network estimation while preserving privacy and security. Moreover, because clients may come from different sources, their data can follow disparate distributions, resulting in heterogeneity. To address these challenges, we introduce a federated personalized learning based method for estimating a Dynamic Bayesian Network from distributed heterogeneous time series data by incorporating a proximal operator as a regularization term. More importantly, our framework has been extended to tensor-like data, such as Electronic Health Records (EHRs). Since the tensor data can be considered as many frontal slices and decomposed into several meaningful components, also known as computational phenotype, we can learn the causal phenotype network without sharing the information of any slice. By leveraging PARAFAC2 decomposition, we introduce a joint-learning approach to infer the causal network. To the best of our knowledge, There is limited work in both research directions. Experimental results on synthetic and real-world datasets demonstrate that both our methods outperform state-of-the-art techniques.

4 - Deepoformer: Deep Operator Learning with Domain-Informed Features for Fatigue Life Prediction**Chenyang Li, New Jersey Institute of Technology, Newark, NJ, United States, Tanmay Kapure, Prokash Roy, Zhengtao Gan, Bo Shen**

Fatigue life characterizes the duration a material can function before failure under specific environmental conditions, and is traditionally assessed using stress-life (S-N) curves. While machine learning and deep learning offer promising results for fatigue life prediction, they face the overfitting challenge because of the small size of fatigue experimental data in specific materials. To address this challenge, we propose, DeepOFormer, by formulating S-N curve prediction as an operator learning problem. DeepOFormer improves the deep operator learning framework with a Transformer-based encoder and a mean L2 relative error loss function. We also consider Stüssi, Weibull, and Pascual and Meeker (PM) features as domain-informed features. These features are motivated by empirical fatigue models. To evaluate the performance of our DeepOFormer, we compare it with different deep learning models and XGBoost on a dataset with 54 S-N curves of aluminum alloys. With seven different aluminum alloys selected for testing, our DeepOFormer achieves an R2 of 0.9515, a mean absolute error of 0.2080, and a mean relative error of 0.5077, significantly outperforming state-of-the-art deep/machine learning methods including DeepONet, TabTransformer, and XGBoost, etc. The results highlight that our DeepOFormer integrating with domain-informed features substantially improves prediction accuracy and generalization capabilities for fatigue life prediction in aluminum alloys.

5 - Jump Gaussian Process Surrogate Modeling with High Dimensional Inputs**Yang Xu, University of Washington, Seattle, WA, United States, Robert Gramacy, Chiwoo Park**

Gaussian process regression (GPR) is a powerful surrogate-modeling tool that delivers principled uncertainty quantification but struggles with piecewise-continuous targets. The Jump Gaussian Process (JGP) addresses this by simultaneously learning local partitions and fitting independent GPs, yet in high dimensions sparse neighborhoods degrade its performance.

We introduce the Locally Mahalanobis Jump Gaussian Process (LM-JGP), a novel method for estimating piecewise continuous regression functions in high-dimensional input spaces. LM-JGP merges the discontinuity detection capability of Jump Gaussian Processes with a built-in local dimensionality reduction mechanism. Our approach employs region-specific projection matrices to capture subspace structure, coupled with latent jump indicators to identify regime changes. We develop a scalable variational inference algorithm that handles both inducing variables and jump indicators, and demonstrate superior prediction accuracy and uncertainty quantification on synthetic and benchmark datasets.

MD62

Building B Level 4 B402

Advancements in Decision-Making for Reliability and Maintenance

Invited Session

Quality, Statistics and Reliability

Chair: Vipul Bansal, University of Wisconsin Madison, Madison, WI, United States

Co-Chair: Akash Deep, Oklahoma State University, Stillwater, OK, United States

1 - "Learning to Recognize the Unseen: A Meta-Learning Approach for Ood Detection in Manufacturing"

Zihan Li, University of Florida, Gainesville, FL, United States

Defect inspection plays a vital role in predictive maintenance for manufacturing systems such as batteries, bearings, and aircraft engines. While neural network (NN)-based methods have demonstrated promising performance in qualification tasks, they often rely on the assumption that the testing system shares the same manufacturing specifications and operational conditions as the training data. In practice, however, systems are frequently heterogeneous, and data availability for the system of interest is limited. This mismatch can significantly degrade model performance and reliability, especially in safety-critical applications.

To address these challenges, we propose a meta-learning-based approach for Out-of-Distribution (OOD) detection in defect inspection. Meta-learning, or “learning to learn,” enables models to quickly adapt to new tasks with minimal data, making it well-suited for heterogeneous manufacturing environments. This is particularly important in modern manufacturing settings where unseen operating conditions, new machinery, or environmental changes may arise. By focusing on OOD detection through meta-learning, our research aims to build more robust and trustworthy predictive models that can generalize beyond the closed-world assumption and maintain reliability under evolving real-world conditions.

2 - Two-Layer Multi-Agent RL for Scalable Fleet Maintenance Under Constraints

Ahmad Salehiyan, Oklahoma State University, Stillwater, OK, United States, Akash Deep

Managing large fleets under constraints on crews, parts, and budgets requires scalable and coordinated maintenance planning. This work presents a two-layer, multi-agent framework for condition-based maintenance. At the lower layer, each asset is modeled as a Markov Decision Process (MDP) that tracks its own degradation and selects maintenance actions. At the upper layer, a central coordinator monitors the fleet, enforces constraints on resource usage and downtime, and provides high-level guidance. Both layers are trained jointly using an actor-critic reinforcement learning approach: the central agent maintains a shared critic, while each asset independently updates its actor. This architecture allows each asset to adapt to its specific condition while ensuring that fleet-level performance remains within budget and availability targets. We demonstrate our framework’s effectiveness with numerical experiments and a real-world fleet maintenance case study.

3 - Condition-Based Maintenance Planning for Partial Observable Multi-Component Systems

Han Hu, University of Wisconsin-Madison, Madison, WI, United States, Vipul Bansal, Shiyu Zhou

Condition-based maintenance (CBM) is a wide-used maintenance planning framework as it makes better-informed maintenance decisions based on system conditions. However, the prevalent existence of partial observability in various systems brings difficulty to CBM problems, especially those of complex systems. This paper addresses the challenge of CBM planning for partially observable large-scale systems with K -out-of- N configuration and economic dependencies. The computational intractability of directly applying Partially Observable Markov Decision Process (POMDP) approaches to such systems motivates our proposed Component-Wise POMDP (CW-POMDP) framework. By decomposing the system-level problem into component-level POMDPs while preserving economic dependencies, we develop efficient maintenance policies using the Successive Approximations of the Reachable Space under Optimal Policies (SARSOP) solver. Theoretical analysis establishes bounds on the optimality gap, while numerical experiments demonstrate that CW-POMDP consistently outperforms benchmark methods across varying system sizes, setup costs, and observation uncertainty levels. A turbofan engine fleet maintenance case study validates our approach’s practical effectiveness for real-world applications.

4 - Towards Failure-Mode-Aware Federated Prognostics: Selective Collaboration for Failure Time Prediction

Yuhan Hu, NC State University, Raleigh, NC, United States, Xiaolei Fang

Federated prognostics enable clients to collaboratively develop failure time prediction models while keeping each client’s data local and confidential. However, existing federated prognostic models face several challenges in real-world applications, particularly due to data heterogeneity across clients and high dimensionality of degradation signals. To address these issues, this talk introduces a three-step federated prognostic framework that facilitates efficient feature extraction and tailored model training while preserving data privacy. We first cluster machines with similar degradation patterns via a LASSO-based approach without centralizing raw data. Then, within each cluster, a federated data fusion method based on robust principal component analysis is applied to extract informative representations of degradation signals. Finally, each cluster utilizes the fused features to train a personalized prognostics model using log-location-scale regression. Numerical studies demonstrate the superior performance of the proposed model compared to existing methods.

5 - Tensor-Efficient High-Dimensional Q-Learning

Junyi Wu, University of Wisconsin, Madison, Madison, WI, United States, Dan Li

High-dimensional reinforcement learning faces challenges with complex calculations and low sample efficiency in large state-action spaces. Q-learning algorithms struggle particularly with the curse of dimensionality, where the number of state-action pairs grows exponentially with problem size. Current solutions include function approximation methods and neural network-based approaches like Deep Q-Networks, but these often suffer from instability and require extensive hyperparameter tuning. We propose Tensor-Efficient Q-Learning (TEQL), which applies low-rank tensor decomposition via block coordinate descent on discretized state-action spaces, incorporating novel exploration and regularization mechanisms. The key innovation is an exploration strategy that combines approximation error with visit count-based upper confidence bound to prioritize actions with high uncertainty, avoiding wasteful random exploration. Additionally, we incorporate a frequency-based penalty term in the objective function to encourage exploration of less-visited state-action pairs and reduce overfitting to frequently visited regions. Empirical results on large discrete control tasks demonstrate that TEQL outperforms conventional matrix-based methods and deep RL approaches in both sample efficiency and total rewards, making it suitable for resource-constrained applications, such as space and healthcare where sampling costs are high.

MD63

Building B Level 4 B403

The Transforming Roles of AI In Economic Theory and Applications

Invited Session

Artificial Intelligence

Chair: Chenglong Zhang, Chinese University of Hong Kong, Shenzhen, Shenzhen, 518172

1 - Data-Driven Policy Learning for Continuous Treatments

Yue Fang, Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Chunrong Ai, Haitian Xie

This paper develops a framework for policy learning with continuous treatments using observational data. Unlike discrete treatment settings, continuous treatments pose greater challenges due to the need for nonparametric estimation of population welfare and the potential infinite-dimensionality of the policy space, often subject to shape constraints as usually considered in economic applications. To address these challenges, we propose an approximation-based approach that embeds the policy space into a sequence of finite-dimensional function classes, allowing scalable optimization and principled generalization. For a given policy, we estimate empirical welfare using kernel-based methods, and we handle both known and machine-learned propensity scores. In the latter case, we adjust the empirical welfare to correct for the bias introduced by estimating the propensity function with flexible machine learning models. Our method selects tuning parameters—such as bandwidths and policy space complexity—via data-driven procedures, enabling automation and adaptivity. We derive oracle inequalities that characterize the regret of the learned policy relative to the optimal one. As a key implementation, we integrate deep neural networks as a policy class to capture complex treatment-outcome relationships, demonstrating the applicability of our approach in high-capacity AI-driven models.

2 - Consumer-Led Data Sharing

Yifan Dou, Fudan University, Shanghai, China, People's Republic of, Xiaoyang Zhang

Serving as the cornerstone of the AI development, consumers’ data is receiving growing attention from firms and regulators. One of the recent legislation efforts in Europe is to enable consumers to initiate data-sharing between firms. This new paradigm introduces economic implications that remain largely unexplored in existing literature.

This paper employs a stylized model in which consumers can mandate the data-rich firm to share the data to the data-poor one. Our results indicate that consumer-led data sharing intensifies competition, which increases the aggregate consumer surplus and social welfare. However, it may also harm certain consumer segments, particularly those who are loyal to the data-poor firm, due to firm-side strategic price adjustments. These findings offer timely insights to understand the economic impact of new advances in the data regulation.

3 - Deep Learning for Double Auction

Chenglong Zhang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Jiayin Liu

Double auctions, where multiple buyers and sellers bid simultaneously, present unique challenges due to incomplete information, complex interactions among traders and the third-party auctioneer platform. While significant progress has been made in the design of single-sided or multi-item auctions, the optimal mechanism design for double auctions remains an open problem. In this paper, we explore the use of deep learning methods to approximate optimal double auction mechanisms in a general theoretical framework. We first identify three key issues for DL-based methods to auction design: scalability to varying market sizes, inefficiencies in enforcing incentive compatibility (IC) and individual rationality (IR), and instability during the training process. To address these issues, we then propose a transformer-based architecture that models market participants as sequences, enabling flexibility across different market sizes. We reformulate and relax the IC and IR constraints into a tractable form and propose a gradient conflict elimination strategy to enhance constraint enforcement during training. Experimental results show that our approach significantly outperforms traditional analytical mechanisms as well as existing deep learning methods designed for simpler auction settings.

4 - Sharing AI Technologies with Third-Party Developers: Instant, Delayed, OR Not?

Yifan Liu, Tianjin University, Tianjin, China, People's Republic of, Jianqing Chen, Minqiang Li, Haiyang Feng

The rapid advancement of Artificial Intelligence (AI) and the increasing availability of data have motivated software platforms to integrate AI technologies into their first-party applications (apps). While platforms earn commission revenue from third-party developers, the competition between first-party and third-party apps creates a strategic dilemma: should the platform share its AI technology with third-party developers, and if so, share it instantly or with a delay? We develop a two-period game-theoretical model in which a platform, offering an AI-powered first-party app that competes with a third-party app, chooses among three strategies: Instant Sharing, Delayed Sharing, and No Sharing. Special attention is paid to data network effects, a distinguishing feature of AI technologies, implying that the quality of AI grows with accumulated user data. We find that although the Delayed Sharing strategy leads to the lowest AI quality in the second period, it can be optimal by mitigating price competition and preserving the first-party app's competitive advantage in the first period, especially when the AI transfer rate is high and the initial AI quality is moderate. Conversely, the Instant Sharing strategy, despite diminishing the first-party app's technological exclusivity, softens price competition in the first period and achieves the highest AI quality in the second period, making it optimal when both the AI transfer rate and the initial AI quality are low. A stronger data network effect has nonmonotonic impacts on the optimality of these strategies: it increases (decreases) the platform's incentive to retain AI when the initial AI quality is high (low).

5 - Impacts of Knowledge-Sharing Communities on Nonprofessional Hosts in Sharing Economy

jiana meng, Iowa State University, Ames, IA, United States, Chen Zhang, He Li

Online communities around sharing economy platforms have emerged as a new type of online community. Gig workers who provide services or share assets through digital platforms use these online communities to facilitate knowledge exchange. Although some studies have explored information exchange among sharing economy participants, they primarily focus on third-party forums such as Facebook and Reddit. Research on peer-to-peer knowledge sharing in sharing economy platform-sponsored online communities remains limited. These communities, which are typically integrated into the sharing platform, are designed to support gig workers by providing updates from the platform and more importantly facilitating peer support and improving their engagement. The content is usually moderated by the platform owner. This study explores how peer-to-peer knowledge sharing in the sharing economy platform-sponsored online community influences nonprofessional hosts' strategies, performance, and exit on the platform. We find that increased peer-to-peer knowledge sharing in Airbnb-hosted online communities contributes to lower price differentiation, improved demand, and decreased host exits. Our results also suggest two mediation effects. Peer-to-peer knowledge sharing in these communities increases demands, which in turn decreases host exits whereas such knowledge sharing reduces price differentiation, which increases demands and ultimately reduces host exits.

MD64

Building B Level 4 B404

Buyer and Suppliers : Models, Agents and Decision Behavior

Invited Session

Behavioral Operations Management

Chair: Brent Moritz, Penn State University, University Park, PA, 16802, United States

1 - Biases of Humans, of AI, and of Humans with AI

Anton Ovchinnikov, Smith School of Business, Queen's University, Kingston, ON, Canada, Yang Chen, Sam Kirshner, Tracy Jenkin, Meena Andiappan

Our research examines whether individuals change their behavioral biases when interacting with LLMs that may also exhibit biases. We design a novel interactive AI experiment and show that AI biases indeed impact human decision-making and discover a new form of nudging. Importantly, we show that how humans interact with AI matters for decision-making.

2 - Distributional Preferences of Large Language Models in Supply Chains: the Roles of Supplier, Buyer, and Relationship

Jie Ning, Case Western Reserve University, Cleveland, OH, United States, Ye Liu

As generative artificial intelligence (GenAI) that harnesses large language models (LLMs) transform the world, more and more companies start to use GenAI in various supply chain interactions such as contract negotiation. This trend makes it imperative to understand the decision-making traits of LLMs, under different personas (supplier vs buyer), and under different supply chain relationships. It has been widely reported that LLMs are more generous and efficiency-driven than the average human in interpersonal contexts. This research takes a first step in characterizing LLMs' distributional preferences in a pricing context in supply chains. We show that GPT-4o has markedly different preferences in a business-to-business setting than an interpersonal setting. In a supply chain, GPT-4o prices products significantly differently when acting as a supplier or a buyer, being consistently more efficiency-driven as a supplier and more fairness-minded as a buyer. GPT-4o may also have surprising perceptions of typical supply chain relationship descriptions. Our findings provide guidance on the customization and design of GenAI tools for companies that currently use, or plan to use, LLMs in supply chain interactions.

3 - Artificial Agents in Operations Management Experiments

Sam Kirshner, UNSW Business School, Sydney, Australia

Built on large language models (LLMs), silicon samples offer a promising approach for advancing behavioral research by simulating decision-making processes and replicating human decision-making in controlled experiments. This study evaluates LLMs' potential to replicate behavioral hypotheses in operations management (OM) by conducting nine studies from the *Management Science Replication Project*. I develop an accessible framework where GPT-4o agents are assigned roles based on System 1 decision-making and participate in the experiments. Remarkably, the silicon samples replicated the core hypotheses in eight of the nine studies, with treatment effects closely matching the direction and magnitude of human decisions. The natural language explanations generated by GPT agents offer insight into their decision logic, revealing whether observed choices align with the behavioral theories being tested. These findings position silicon samples as a valuable tool for replication research, helping researchers distinguish between sampling limitations and theoretical divergence. More broadly, the results support using silicon samples to scale replication, design interventions, and conduct high-fidelity pretests before lab or field deployment.

4 - Supplier Selection and Capacity Constraints

Brent Moritz, Penn State University, University Park, PA, United States

We investigate optimal and observed sourcing decisions when a buyer faces capacity constraints, and suppliers differ in their reliabilities. We consider two suppliers, where one supplier is more costly yet reliable, and the other supplier is less expensive but unreliable. Analytically, we demonstrate that a firm's capacity constraints should impact their sourcing decisions though sunk costs should not impact the decision. Capacity constraints generally favor sourcing more inventory from the reliable supplier. We evaluate these findings in a behavioral experiment. In contrast to theoretical and optimal predictions, when capacity is constrained, we observe that decision makers tend to source more from the unreliable supplier. Similarly, when sunk costs are high, decision makers sourced more from the reliable supplier.

5 - Examining the Effect of Supply Chain CSR Performance and Price Factors on Consumers' Purchasing Decisions

Tim Kraft, North Carolina State University, Raleigh, NC, United States, Junhao Yu, Rob Handfield, Rejaul Hasan, Marguerite Moore

We examine consumers' evaluations of a retailer's communication of its manufacturer's CSR performance (both positive and less than perfect). Using an online shopping context, we conduct both a non-incentivized study and an incentivized study. In both studies, participants were presented with a purchase scenario featuring an apparel item with a CSR performance label and answered questions assessing their purchase intentions. In the incentivized study, a participant's purchase decision influenced the composition of a prize package they could receive (which may include a T-shirt depending on their decision). First, we assess whether communicating CSR performance impacts consumers' purchase intentions. Second, we explore potential mediation paths of this effect in order to illustrate the mechanisms (price fairness and product desirability) by which CSR communication impacts consumers' purchase intentions. Finally, we examine how two price dimensions, retail price and historical price paid, moderate these mediation effects.

We find evidence of our hypothesized indirect effects at the aggregate level under different CSR performance levels, with price fairness being a particularly robust mediator across multiple CSR performance levels. Our moderated mediation study further reveals that the strength of the indirect effects via price fairness varies with consumers' historical price paid and retail price. Specifically, historical price paid moderates the impact of CSR performance on price fairness, with positive (negative) indirect effects evident for consumers with a relatively lower (higher) historical price paid. Conversely, retail price moderates the ancillary impact of price fairness on product desirability, amplifying the moderation effect of historical price paid.

MD65

Building B Level 4 B405

Adaptive experimentation and LLM in Business

Invited Session

Artificial Intelligence

Chair: Wenjia Ba, University of British Columbia, Vancouver, BC, Canada

Co-Chair: Zikun Ye, University of Washington, Seattle, WA, United States

1 - Understanding the Impact of Sampling Quality in Direct Preference Optimization

Yumo Bai, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States

We study the role of the sampling distribution in Direct Preference Optimization 2 (DPO) and aim to understand its impact on DPO's training dynamics. Our analyses 3 show that both the solution space and the convergence behavior of DPO depend 4 on the support and quality of the generating distribution. We first analyze how 5 distribution of responses influences policy updates during gradient descent, drawing 6 connections to common phenomena found in practice. We then design a simplified 7 yet well-structured alignment model as a proxy, and develop quantitative results 8 showing how more frequent high-quality responses amplify the gradient signal 9 and improve the optimization landscape, leading to more effective policy learning. 10 Our theoretical findings are supported by empirical experiments and provide a 11 principled justification for the online DPO framework in practice.

2 - Orlm: a Customizable Framework in Training Large Models for Automated Optimization Modeling

Chenyu Huang, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, Zhengyang Tang, Shixi Hu, Ruoqing Jiang, Xin Zheng, DongDong Ge, Benyou Wang, Zizhuo Wang

Optimization modeling plays a critical role in the application of Operations Research (OR) tools to address real-world problems, yet they pose challenges and require extensive expertise from OR experts. With the advent of large language models (LLMs), new opportunities have emerged to streamline and automate such task. However, current research predominantly relies on closed-source LLMs such as GPT-4, along with extensive prompt engineering techniques. This reliance stems from the scarcity of high-quality training datasets for optimization modeling, resulting in elevated costs, prolonged processing times, and privacy concerns. To address these challenges, our work is the first to propose a viable path for training open-source LLMs that are capable of optimization modeling and developing solver codes, eventually leading to a superior ability for automating optimization modeling and solving. Particularly, we design the *OR-Instruct*, a semi-automated data synthesis framework for optimization modeling that enables customizable enhancements for specific scenarios or model types. This work also introduces IndustryOR, the first industrial benchmark for evaluating LLMs in solving practical OR problems. We train several 7B-scale open-source LLMs using synthesized data (dubbed ORLMs), which exhibit significantly enhanced optimization modeling capabilities, achieving competitive performance across the NL4OPT, MAMO, and IndustryOR benchmarks. Additionally, our experiments highlight the potential of scaling law and reinforcement learning to further enhance the performance of ORLMs. The workflows and human-machine interaction paradigms of ORLMs in practical industrial applications are also discussed in the paper.

3 - Document Valuation in LLM Summaries: a Cluster Shapley Approach

Zikun Ye, University of Washington, Seattle, WA, United States, Hema Yoganarasimhan

Large Language Models (LLMs) combined with the Retrieval-Augmented Generation (RAG) framework are transforming consumer search and information consumption. Consumers increasingly rely on LLM-based summaries (generated from multiple source documents in response to search queries) instead of reading individual documents/sources. While this simplifies the search and information retrieval process, there exists no established method to quantify the contribution of each document to the overall quality of the LLM-summary, and compensate document creators. In this paper, we first propose a theoretical solution to this problem based on the Shapley value framework that is fair and transparent. While this framework is theoretically appealing, it is practically infeasible since the computational costs of calculating Shapley values grow exponentially with the number of documents. We therefore develop a novel Cluster Shapley algorithm, that leverages embedding-based semantic similarity to cluster similar documents, thereby reducing the number of LLM summarizations and evaluations while maintaining high attribution accuracy. Our approach is particularly well-suited to LLM settings since LLMs themselves can be used to obtain high-quality text embeddings. We apply our approach to a large dataset consisting of Amazon reviews, where LLMs are used to summarize review content. Using a series of numerical experiments, we show that our algorithm achieves up to a 40% reduction in computation time compared to exact Shapley values, and outperforms other baseline approximations. To our knowledge, this work is the first to propose an economic and computational framework for document value attribution in LLM-summarization systems.

4 - Beyond Automation: Large-Model-Powered Automl Reasoning with Knowledge-Driven Insights (Lark) in Marketing Analytics

Lin Lei, University of Maryland, Robert H. Smith School of Business, College Park, MD, United States, Chenyu Huang, Wenzhe Mao, Dongdong Ge, P. K. Kannan

Marketing organizations often struggle with knowledge retention and consistent decision-making due to rapid employee turnover, shifting markets, and complex analytical demands. This study introduces LARK (Large-Model-Powered AutoML Reasoning with Knowledge-Driven Insights), an AI-based approach specifically designed to enhance marketing analytics effectiveness. LARK integrates AutoML and LLMs to iteratively generate contextually meaningful features for time-series prediction or panel regression tasks, with the access to historical information from similar tasks (from organizational memory). LARK efficiently captures and applies insights from historical marketing analytics projects, significantly improving predictive accuracy, reducing analytical redundancy, and supporting agile strategic decisions. We use sales prediction of Chow Tai Fook, a jewelry brand as the empirical validation to confirm LARK's ability. We also apply the manufacturing regression task of FastPrint, a PCB company, to justify the robust performance of LARK even with limited prior knowledge. Additionally, we put together six different empirical analysis and utilize four

different LLMs to highlight LARK's potential to provide enduring competitive advantages through enhanced organizational memory and analytics agility for marketing teams, especially compared to human experts and standard open-source end-to-end LLMs (TimeGPT, TSFM, etc.)

5 - Online Conditional Value-at-Risk Control

Catherine Chen, Stanford University, Stanford, CA, United States, Lihua Lei

We present a generalization of conformal tail risk control to the Conditional Value-at-Risk (CVaR) in an online setting, enabling principled risk-sensitive prediction for a broad class of black-box learning and generative models. Our approach is designed for non-stationary online settings, where the data distribution may drift over time or even shift adversarially, posing significant challenges for traditional risk control techniques. By drawing on deep connections between conformal tail risk control, adversarial online learning, and the variational representation of CVaR introduced by Rockafellar and Uryasev, we develop a novel framework for online CVaR control. This procedure operates model-agnostically, making it broadly applicable in practice. We validate the performance of our method on a range of real-world datasets, including high-stakes financial forecasting tasks and large language models (LLMs), where tail risks can be particularly costly. Our experiments demonstrate the consistency of our proposed algorithm for maintaining desired risk levels dynamically.

MD66

Building B Level 4 B406

Critical Thinking in the Age of ChatGPT: Insights from Teaching, Trials, and Tools

Invited Session

INFORMS Committee on Teaching and Learning

Co-Chair: Gabriela Gongora-Svartzman, Khoury College, Northeastern University, Miami, FL, United States

1 - Student AI Usage in Higher Education

Daniel Nielubowicz, University of Washington, Bothell, WA, United States

This study investigates the utilization of AI tools by students in higher education, examining usage trends, academic impact, and ethical considerations. It identifies challenges such as digital literacy gaps and access disparities, providing valuable insights for institutions to foster responsible and effective AI integration in academic settings.

2 - The Impact of Generative AI on Student Success: Evidence from a Randomized Controlled Trial in Business Education

Insung Hwang, Columbus State University, Columbus, GA, United States, John Finley

This randomized controlled trial investigates the impact of generative artificial intelligence (AI) tools on student academic success in higher education. The study involves 30 classes across multiple disciplines at a public university, randomly assigned to either an intervention group with structured access to a generative AI assistant (e.g., ChatGPT) or a control group with AI prohibition policies and measures. Over a semester, the intervention group is allowed to use the AI tool for assignments including idea generation, essay drafting, and coding support, with training on ethical and effective use. Key outcomes of the investigation include course grades, assignment quality, student retention in the current and next semester, and the students' job-market-related metrics such as internship placement, interview invitations, and job placement. In addition to estimating the causal effects of generative AI on academic and career-related metrics, the study also explores the boundary conditions across contextual factors and individual factors—such as disciplinary type, course type, instructor approach, and student characteristics (previous academic history/demographic factors)—that moderate the effectiveness of AI integration. This study provides early causal evidence on whether and when generative AI tools can enhance student performance and engagement in academic settings. The findings aim to inform the pedagogical potential of AI and support the development of evidence-based policies and guidelines for its integration in educational settings.

3 - Improving Student's Critical Thinking Skills Using ChatGPT Generated Essays

Zoe Long, University of Florida, Gainesville, FL, United States, Elif Akcali

With the increasing growth of large language models use in higher education, it is increasingly important that students understand the limitations of these tools. To better understand students' current opinions and how they may be affected by being exposed to mistakes made by ChatGPT, we designed this research study. We provided undergraduate industrial and systems engineering students in a supply chain course with an assignment to ask ChatGPT to generate two different essays about two different supply chain topics previously discussed in the course. Students were then asked to identify factual and logical errors within these essays along with incorrect or incomplete citations. Students completed pre- and post-surveys using 5-point Likert scales assessing their agreement with statements about ChatGPT's use in education (1=strongly disagree, 5=strongly agree) and its perceived helpfulness for specific tasks (1=completely unhelpful, 5=extremely helpful). 38 of the 55 students in the course consented to having their data included in this study. We compared the results of the pre- and post-surveys to assess changes in student opinions of ChatGPT. We also investigated whether the number and type of mistakes identified by the students relate to the changes in their opinions. This research will examine what changes were found in students' opinions after being exposed to mistakes made by ChatGPT.

4 - Harnessing the Power of Learning Analytics for Metacognition Skills Development

Anita Lee-Post, University of Kentucky, Lexington, KY, United States, Kun Huang

Metacognition skills are essential for individuals to take control of their learning and thinking, enabling them to achieve both professional and personal goals. When powered by learning analytics, individuals can enhance their metacognitive abilities based on reliable evidence and insights gleaned from data collected about their learning and thinking processes as they interact with the learning management systems to complete learning tasks and activities. In this talk, we will present a research study designed to improve students' academic performance with the use of learning support exercises for the four phases of metacognition: planning, monitoring, evaluation, and reflection. The learning support exercises are labeled as study plan set up, study plan follow up, study plan tune up, and study plan wrap up to represent each of the four phases of metacognition respectively. All exercises involve helping learners align their learning goals with a solid study plan. Results from implementing the research study in an undergraduate Business Analytics online class are discussed.

MD67

Building B Level 4 B407

Data-Enabled Analytics: Integrating Efficiency, Intelligence, and Impact

Invited Session

Data Mining

Chair: Dariush Khezrimotlagh, Pennsylvania State University - Harrisburg, Middletown, PA, 17057, United States

Co-Chair: leili Soltanisehat, University of Massachusetts Dartmouth, North Dartmouth, MA, United States

1 - Beyond the Frontier: Learning from Observed Units in Multi-Stage System Performance Analysis

Dariush Khezrimotlagh, Pennsylvania State University - Harrisburg, Middletown, PA, United States, Yao Chen

This study introduces a practical method for evaluating multi-stage business processes using observed data. Unlike traditional models that rely on hypothetical scenarios, our approach ensures transparency and realistic benchmarking. It empowers decision-makers with reliable insights to assess efficiency, optimize operations, and support data-driven performance improvements.

2 - Enhancing Deep Learning with Data Envelopment Analysis: Market-Aware Recommendation System for Non-Fungible Tokens

Kexin Lin, Worcester Polytechnic Institute, Worcester, MA, United States, Joe Zhu

Non-Fungible Tokens (NFTs), unique digital assets authenticated via blockchain, have emerged as transformative instruments for representing ownership across diverse sectors such as art, gaming, and decentralized finance. The rapid expansion of NFT market, however, brings critical challenges in establishing NFT recommendation

systems that have the ability to reconcile artistic appeal with financial value, extract insight from sparse and noisy interaction data, and match individual preferences to market performance. To tackle these challenges and help users to decide which NFT to buy, we develop a purchase-oriented Transformer Network-based recommendation system that first uses data envelopment analysis (DEA) to assess how NFTs convert intrinsic characteristics into market success. DEA generates market performance scores and establishes reference relationships among NFTs. Using DEA scores with NFTs' intrinsic attributes, we construct three complementary graphs to capture multifaceted relationships: a directed DEA-reference graph that links each NFT to its benchmarks, an undirected performance-cluster graph that connects NFTs in the same performance cluster, and an undirected trait-similarity graph linking NFTs sharing similar traits. A temporal Transformer leverages these relation-specific graphs and DEA scores to recommend NFTs for each user to purchase that simultaneously align with user preferences and exhibit strong market performance. Applications on CryptoPunks dataset demonstrate that our proposed DEA-enhanced Transformer Network (DETN) recommendation system achieves higher accuracy than state-of-the-art baseline models such as collaborative filtering, graph-based methods, and Transformer-based models. By integrating DEA with deep learning method, DETN enhances recommendation precision and supports better informed decision-making processes.

3 - Unified Modeling of Sequential and Cross-Modal Dependencies via Adaptive Multimodal Deep Architectures

Yeganeh Abodollahinejad, Pennsylvania State University, Harrisburg, PA, United States, Dariush Khezrimotlagh

Deep learning systems deployed in real-world settings must often interpret data that is both multimodal and temporally structured. A modular and adaptive architecture is introduced to unify the modeling of sequential and cross-modal dependencies within a single deep learning framework. By combining attention-based temporal reasoning with dynamic cross-modal fusion, the model captures both time-evolving signals and interactions among heterogeneous data sources. Unlike traditional pipelines that treat modalities or time separately, this approach enables unified representation learning, enhancing generalization and reducing reliance on task-specific tuning. The system dynamically adapts to missing data, shifts in modality quality, and variability in sequences through gated components and learnable weighting mechanisms. Multi-objective optimization is employed to jointly balance prediction accuracy, modality confidence, and temporal consistency. Interpretability is supported through modular design and attention-based mechanisms, enabling traceable information flow across inputs and components. The system is designed to remain robust under noisy, imbalanced, and dynamic data conditions. By integrating multiple learning objectives into a cohesive framework, the system provides a principled solution to the challenges of multimodal, sequential inference. Its design supports both rigorous modeling and practical, real-world utility.

4 - A Data-Driven Analysis of ESG Impacts on Financial Performance Across Industries and Regions

leili Soltanisehat, University of Massachusetts Dartmouth, North Dartmouth, MA, United States

This study examines how firms' financial performance is influenced by Environmental, Social, and Governance (ESG) practices, with an emphasis on the moderating roles of industry characteristics and regional contexts. While prior research often focuses on company-level factors, this study addresses the gap by integrating firm- industry- and region-specific dimensions. To capture complex and context-dependent relationships, the analysis applies data-driven models, including various machine learning techniques, enabling nuanced assessments of ESG impacts across sectors and regions. The findings reveal significant heterogeneity in ESG's financial effects, offering actionable insights for investors and managers to adopt tailored ESG strategies aligned with industry dynamics and regional institutional settings.

MD68

Building B Level 4 B408

Advances in Trustworthy Machine Learning: Interpretability, Fairness, and Intelligent Control in Healthcare and Beyond

Invited Session

Data Mining

Chair: Rachel Bennett, University of Oklahoma, Oklahoma City, OK, United States

1 - Robust Outcome Optimization Through Multi-Model Inverse Classification

Shun Dong, University of Kansas, Lawrence, KS, United States, Michael Lash, Qihang Lin, Nick Street

Predictive models have become central to modern decision-making across a wide range of domains, including finance, healthcare, and manufacturing. While these models excel at forecasting future outcomes, they often fall short in offering actionable guidance. Inverse classification, a key technique within the broader framework of algorithmic recourse, addresses this limitation by generating prescriptive recommendations alongside predictions. In this work, we introduce a multi-model inverse classification framework that leverages a novel proximal linear algorithm designed to solve difference-of-weakly-convex optimization problems. Our method enables decision-makers to steer outcomes toward desirable targets while explicitly accounting for a defined level of "risk." We evaluate the proposed framework on real-world datasets, and the empirical results demonstrate its effectiveness and robustness in generating outcome-optimizing prescriptions, thereby validating its practical applicability.

2 - Smarter Cells, Smarter Control: Reinforcement Learning for Fermentation Optimization

Airi Shimamura, University of Oklahoma, Norman, OK, United States, Farid Talebnia Rowshan, M. Soheil Hemmati, Talayeh Razzaghi

Optimizing bioprocesses is a key challenge in biomanufacturing, with fermentation playing a pivotal role in determining both productivity and profitability. However, the inherent complexity of cellular mechanisms poses significant challenges for developing accurate mechanistic models that capture these dynamics. Traditional optimization approaches often rely on extensive experimental datasets, making them costly and time-consuming. To address these limitations, we propose a reinforcement learning (RL) framework that formulates the bioprocess as a sequential decision-making problem. RL is particularly well-suited for this application, as it learns optimal control policies through iterative interaction with the environment, without requiring explicit system models. The proposed RL-based control strategy is designed to optimize fermentation conditions by maximizing protein production while minimizing impurity accumulation and operational expenditures. The results demonstrate the potential of RL as a data-efficient, adaptive optimization tool for complex biomanufacturing processes.

3 - Decision Support Systems for Health Practitioners: An Interpretable Deep Learning Framework to Navigate Diagnostic Uncertainty

Maryam Kheirandish, Emory University, Atlanta, GA, United States, Shengfan Zhang, Ryan Suk

Health systems are currently facing unprecedented financial challenges, exacerbated by the suboptimal utilization of diagnostic tests. Both the overuse and underuse of these tests deplete essential health resources and impede timely diagnosis and effective treatment. A primary driver of test overuse is the inherent uncertainty in medical decision-making; merely ordering additional diagnostic tests does not guarantee the prevention of missed conditions. To counteract this, we propose an interpretable deep learning framework designed to assist physicians by providing clear insights into a patient's current health status, leveraging existing test results and associated uncertainties. This empowers physicians to make informed decisions regarding the necessity of additional tests or the feasibility of proceeding with treatment based on available data, thereby ensuring conditions are not overlooked. While predictive uncertainty may suffice for decision-making at the diagnosis stage, ongoing treatment monitoring necessitates a more dynamic, sequential decision-making approach. In this study, we introduce an interpretable deep learning framework that integrates uncertainty quantification within outcome prediction models to significantly improve treatment monitoring strategies. This framework is specifically designed to identify scenarios where additional monitoring tests can enhance the predictive certainty of treatment outcomes. The model considers the immediate impact of test ordering on current predictive uncertainty, as well as its influence on subsequent predictions used in future patient follow-ups. We demonstrate the practical utility and efficacy of this model through its application to enhance the treatment monitoring of tuberculosis patients.

4 - Improving Maternal Health: Predicting Length of Stay in Preeclampsia with Imbalanced Learning

Talayeh Razzaghi, University of Oklahoma, Norman, OK, United States, Leama Tah, Rachel Bennett, Mulla D Zuber

Hypertensive disorders during pregnancy, including preeclampsia, are among the leading causes of maternal and neonatal morbidity and mortality. Preeclampsia affects approximately 2-8% of pregnancies in the United States, with a disproportionately higher incidence among African American women (6.04%) compared to Caucasian women (3.75%). The condition often requires intensive care unit (ICU) admission, leading to prolonged hospital stays and increased healthcare burdens. This study aims to predict the length of stay for preeclamptic patients using supervised machine learning techniques applied to an imbalanced healthcare dataset. We exploit and compare several approaches to address imbalanced data to improve predictive performance and model trustworthiness.

5 - Survival Analysis for Risk Stratification of Clinical Events in Heart Failure Patients Using Longitudinal EHR Data

Maryam Jafaripakzad, University of South Florida, Tampa, FL, United States, Jorge Acuna, Jose Zayas Castro, Joel Fernandez

Heart failure is a chronic condition associated with frequent hospitalizations and a high risk of adverse clinical events, posing significant challenges to both patients and healthcare systems. Using time-dependent data collected in real-world clinical settings can improve our ability to anticipate patient deterioration and guide timely interventions. In this study, we develop a deep learning framework that utilizes longitudinal electronic health record data extracted from the EPIC system to model patient trajectories and predict critical clinical events in the months ahead. We propose a hybrid model combining a Temporal Convolutional Network (TCN) with an autoencoder to capture the temporal dynamics and latent characteristics of individual patients. The autoencoder component extracts compact, informative representations of multivariate patient data, while the TCN captures sequential patterns over time. This model enables survival analysis of multiple adverse events such as cardiogenic shock, transplant, LVAD implantation, and mortality. Our dataset includes a diverse cohort of heart failure patients with varying numbers of hospital or clinical encounters, irregular visit intervals, and heterogeneous clinical measurements which reflect the complexity of our dataset. The model is trained to predict the risk of each event in the upcoming months which offers a personalized risk profile for each patient. This approach not only improves predictive performance but also allows for interpretable patient representations that can support risk stratification and care planning.

MD69

Building B Level 4 B409

Innovation and Technology Adoption

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Niyazi Taneri, University of Cambridge, Cambridge, United Kingdom

2 - Improving Farm Productivity and Technology Adoption Through LLM-Powered Digital Advisory: a Randomized Controlled Trial Among 2,000 Smallholder Rice Farmers

Canberk Ucel, ESSEC Business School, Paris, France, Ali Aouad, Alp Sungu, Niyazi Taneri, El Ghali Zerhouni

Low adoption and ineffective implementation of sustainable agricultural technologies continue to undermine productivity and environmental outcomes in smallholder farming systems. This study presents one of the first large-scale randomized controlled trials (RCTs) testing the use of Large Language Model (LLM)-based AI advisory bots to address these challenges. We examine whether delivering personalized, on-demand guidance via common messaging platforms on farmers' smartphones can improve farm management practices and the adoption of Direct-Seeded Rice (DSR) technology—a practice with demonstrated environmental and economic benefits. Our ongoing experiment targets 2,000 smallholder rice farmers across four major rice-growing regions in the Philippines. Participants are randomly assigned to a control group or one of two treatment arms: access to an AI advisor bot, or access with additional incentives designed to encourage initial engagement. The advisory bot integrates technical content from Bayer CropScience and public sources and is tailored to local farming contexts. We collect detailed survey and interaction data throughout the growing season and apply econometric and qualitative methods to evaluate technology adoption, farmer engagement, productivity, and AI-human interaction patterns. The study advances research at the intersection of agricultural development, operations management, and AI deployment in low-resource environments. It offers practical insights into the design of scalable digital advisory systems and the behavioral dynamics underpinning sustained technology use. Findings will inform policy and industry efforts aimed at improving agricultural productivity, climate resilience, and the cost-effectiveness of extension services in developing countries.

3 - Incentivizing Project Managers Competing for R&D Funds

Niyazi Taneri, University of Cambridge, Cambridge, United Kingdom, Ozge Tuncel, Nektarios Oraopoulos, Jochen Schlapp

Theory suggests combining shared and individual incentives can align the incentives of managers competing for R&D funds. We characterize the optimal contract that incentivizes project managers to evaluate their projects and truthfully report evaluation outcomes, which determines the firm's resource allocation decision. Through a series of treatments, we then explore the contract's interaction with different business environments. Our experimental results indicate that the optimal contract alone doesn't sufficiently motivate project managers, suggesting that the business environment, paired with the optimal contract, plays a crucial role in moving toward theoretical outcomes.

4 - Do Workers and Customers Benefit from Competition Between on-Demand Service Platforms

Sangyoon Cha, University of Michigan Ann Arbor, Ann Arbor, MI, United States, Saif Benjaafar, Shihong Xiao, Xiaotang Yang

We study competition between two on-demand service platforms: an incumbent and a new entrant. The platforms compete for both workers and customers, deciding on wages to pay workers and prices to charge customers. Workers, who are heterogeneous in their opportunity costs, are sensitive to both the wage and utilization. Customers, who are heterogeneous in their preferences for the platforms, are sensitive to both the price and delay. We show that there are cases where both workers and customers are worse off after the new platform enters the market (with workers being busier but earning less, and customers paying more but experiencing more delay). Moreover, prohibiting workers from multihoming (i.e., working for both platforms) can flip the outcomes for workers (they become less busy but earn more) while customers continue to be worse off (they pay more).

Monday, October 27, 4:15 PM - 5:30 PM

ME01

Building A Level 3 A301

FAMU-FSU College of Engineering/Princeton Consultants

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - At the Intersection of AI, Data, and Advanced Materials and Manufacturing

Changchun Zeng, Florida State University, Tallahassee, FL, United States

This presentation overviews the diverse research in the Department of Industrial and Manufacturing Engineering (IME). We have an outstanding group of faculty members working in the areas vital to the national interest and to the betterment of society, i.e., advanced materials and manufacturing, nanomaterials and nanotechnology, smart materials, data science and analytics, artificial intelligence, machine learning and deep learning, systems engineering, transportation systems and supply chain logistics. Our faculty members are involved in these exciting areas on projects sponsored by prestigious funding agencies, i.e., NSF, NIH, DOE, DOD, NASA, AFRL, ONR, to name a few.

2 - Optimization Solution Development and Deployment: a Framework for Success

Irv Lustig, Princeton Consultants, Coral Gables, FL, United States

For over 44 years, Princeton Consultants has developed and deployed numerous applications that leverage optimization to make better decisions for our clients. We have developed a set of best practices over that time that are also reflected in the recently published INFORMS Analytics Framework (IAF), for which Irv was a key contributor. The IAF provides a roadmap for successful analytics projects, including optimization, and contains specific tasks related to risk management. Irv will describe

both the IAF and the Princeton 20, a set of 10 environmental and 10 technical risks to be evaluated at the commencement of any project. Understanding these risks in the context of the framework leads to best practices that improve the chances of success for your next optimization project.

ME02

Building A Level 3 A302

2025 INFORMS Prize

Award Session

INFORMS Prize 2025

Chair: Christy Blevins, INFORMS, Catonsville, MD, United States

1 - Inform's Prize-Micron

Sanjay Dasam, Micron Technology, Eagle, ID, United States, Shiva Esturi

Micron Technology, a global leader in memory and storage solutions, is dedicated to transforming the way the world uses information to enrich life for all. Founded in 1978 in Boise, Idaho, Micron has grown to over 50,000 employees worldwide, generating \$25.1 billion in revenue in fiscal year 2024. Micron's industry-leading technologies enable the world around us – from cutting edge artificial intelligence solutions to automotive, Industrial IoT, mobile, and personal computing. At Micron, we leverage a variety of operations research (O.R.) and analytics applications to optimize our operations and decision-making processes. Our initiatives have significantly improved planning, procurement, manufacturing, and logistics and fulfillment. For instance, Micron's use of hierarchical linear programming to optimize supply planning has improved supply-constrained demand satisfaction by 4%, reduced over \$50M in working capital, and recurring manufacturing cost reduction of more than %15M. Micron Master Production Scheduler enhances production schedules for our Wafer Fabrication using advanced analytics. This tool has achieved a wafer fabrication throughput increase of 0.5%, equivalent to \$288 million. Additionally, our Capacity Planning Optimization solution uses a combination of mathematical and simulation techniques to generate optimal capacity plans, helping Micron save more than \$100 million every year. Micron's innovative use of O.R., data science, machine learning, and analytics has not only enhanced our operational efficiency but also positioned us as a leader in the semiconductor industry. We are proud to share our journey and achievements with the INFORMS community.

ME03

Building A Level 3 A303

Artificial Intelligence and Future of Work

Invited Session

eBusiness

Chair: Guohou Shan, Boston, MA, United States

1 - Charting the Unseen: Uncovering the Dynamic Processes Underlying Human-Genai Interaction in Complex Problem Solving

Ziyi Zhao, Temple University, Philadelphia, PA, United States, Kanghyun Cho, Aleks Aaltonen

Generative artificial intelligence (GenAI) is reshaping the landscape of work by redefining how professionals solve complex problems, moving beyond traditional search methods to generate innovative and context-specific solutions. This transformative shift has a profound impact on job roles, responsibilities, and the skills required for effective human-AI collaboration. IT professionals face the critical task of navigating these changes, balancing the potential productivity and innovation gains of GenAI adoption with the need to manage technological risks and ethical considerations. This study directly addresses how IT professionals evaluate and use GenAI-generated solutions in complex problem-solving tasks. Specifically, this research introduces Screen Tracking of Research Engagement, an innovative methodology that employs continuous video-based monitoring of participant interactions within the research environment. By capturing detailed clickstreams, navigation paths, and interaction patterns, this method provides rich contextual insights into participants' cognitive processes and decision-making strategies. The analysis of thousands of minutes of videos of dynamic interactions of GenAI illustrates different strategies that professionals employ when interacting with GenAI. As a result, this research underscores the importance of cultivating prompt representation technique that enables professionals to iteratively guide GenAI toward accurate and contextually tailored outputs.

2 - From Error to Inspiration: Exploring the Impact of AI Hallucination on Creativity

Fanshu Li, George Mason University, Fairfax, VA, United States, SIDDHARTH BHATTACHARYA, Bowen Lou

Abstract

As AI tools like large language models (LLMs) become more embedded in creative work, it is important to understand how they affect human idea generation. One of the most intriguing aspects of these LLMs is their tendency to produce hallucinated content. While such content can mislead or confuse users, it can also prompt unexpected connections and spark original ideas. This study explores the complex role of AI hallucination in shaping human creativity. Rather than treating inaccuracies as purely negative, we examine how different levels of hallucination may hinder or enhance divergent thinking. Theoretically, this research contributes to creativity studies by introducing the idea that hallucinations can inspire novel thinking, challenging the traditional assumption that errors only interfere with performance. It also adds to human-AI interaction research by showing how users engage with imperfect systems during open-ended tasks. Practically, the findings offer guidance for designing AI tools that better support creativity. By identifying conditions under which hallucinations are helpful or harmful, developers and users can adjust system settings and workflows to make the most of AI assistance. These insights are especially relevant for creative industries, education, and ethical AI deployment.

3 - The Double-Edged Sword of AI Assistants in Customer Service: Balancing Efficiency and Human Skill Retention

Shizhen Jia, Quinnipiac University, Hamden, CT, United States, Guohou Shan

The increasing adoption of artificial intelligence (AI) assistants in customer service has transformed how organizations manage customer interactions, offering benefits such as enhanced efficiency, cost reduction, and improved response times. However, as AI systems take on routine tasks, concerns arise regarding their long-term impact on human employees' skills, job satisfaction, and autonomy. While existing research has explored AI's role in augmenting human capabilities and improving customer satisfaction, little attention has been paid to the risks of overreliance on these technologies.

This study addresses this gap by conducting a longitudinal field experiment in a large company's customer service department. Three hundred representatives are randomly assigned to an AI-assisted group or a control group using traditional methods. Over six months, performance metrics—including response time, resolution rates, and customer satisfaction—are tracked alongside employee problem-solving skills, job satisfaction, and autonomy. The study also introduces an intervention, incorporating "AI-free" days and gamified challenges to mitigate potential skill erosion and promote active engagement.

By examining the trade-offs between AI-driven efficiency gains and the long-term implications for human workers, this research provides empirical insights into the evolving role of AI in customer service. Findings will contribute to the discourse on responsible AI adoption and inform strategies for balancing technological advancement with human skill development.

4 - AI's Job Shakeup: Analyzing the Uneven Impact of Ai Adoption on Labor Demand

Xiupeng Wang, Northeastern University, Boston, MA, United States, Bin Gu, Mohammad Soltanieh-ha

This research investigates the influence of Artificial Intelligence (AI) adoption on labor demand within publicly traded enterprises. We develop A novel metric of AI adoption, derived from the analysis of 10-K filings, and AI activities are classified into three distinct categories: "Ops" (operational use), "Dev" (AI development), and

“None.” This classification is conducted using an unsupervised methodology leveraging OpenAI’s GPT-4o model, guided by predefined category prompts and is trained with sentence transformer “all-MiniLM-L6-v2” model. Utilizing AI exposure measures at the occupational level from Webb (2019), findings indicate a negative correlation between AI adoption and job postings in occupations with high impact across the entire sample, whereas no significant correlation is observed with low impacted occupations. Heterogeneity analyses reveal that these effects are predominantly attributed to non-IT service sectors and firms lacking internal AI technology development. This study demonstrates the uneven association of AI with labor demand, highlighting disparate levels of disruption across specific occupations and industries.

5 - Generative AI May be Good for Artists: Understanding the Impact of Artistic Style on Preference and Willingness to Pay

J. Bell, Penn State University, Phoenix, AZ, United States, Wen Wang, Jeffrey Dotson, David Schweidel

Generative AI tools, such as MidJourney, Stable Diffusion, and DALL-E, allow users to generate high-quality images using simple text prompts and have seen wide adoption.

However, the ability of these tools to mimic the styles of specific artists has raised ethical and economic concerns. Artists argue that their intellectual property is being exploited without consent or compensation, fueling calls for fair remuneration. In this paper, we propose a constructive alternative: generative AI as a means for artists to monetize their unique styles, transforming a perceived threat into an economic opportunity. Using a multi-method approach, including large-scale prompt analysis, deep learning models, and conjoint choice experiments, we demonstrate that artist styles can enhance image quality, increase consumer preference, and drive higher willingness to pay (WTP) even under conditions of incentive alignment. Importantly, consumers express strong support for fair compensation schemes, creating a viable pathway for artists to participate meaningfully in the generative AI ecosystem. These findings have implications for technology developers, artists, and policymakers seeking to balance innovation with fairness.

ME04

Building A Level 3 A304

Asset Pricing and Market Behavior

Contributed Session

Contributed

Chair: Ruobing Yan, Harbin Institute of Technology, Harbin, N/A, China, People's Republic of

1 - Hybrid Hidden Markov Models for Modeling Equity Return Dynamics

Abdulrahman Alswaidan, Cornell University, Ithaca, NY, United States, Jeffery Varner

This talk introduces a hybrid probabilistic framework for modeling equity returns, designed to capture empirical phenomena such as heavy tails, volatility clustering, and regime-switching behavior commonly observed in financial markets. Standard models, such as geometric Brownian motion, rely on assumptions of normality and constant volatility, which often fail to reflect the complex dynamics of real-world asset returns. Our approach integrates discrete-state Hidden Markov Models (HMMs) with a Poisson jump component, enabling the simulation of both persistent market regimes and abrupt return shocks. Each hidden state represents a distinct market condition, ranging from severely bearish to strongly bullish with transitions governed by empirically estimated Markov chains. The Poisson process captures rare but significant return jumps, enhancing the model’s ability to replicate observed return distributions. We validate the framework using a dataset of over 400 U.S.-listed equities and exchange-traded funds (ETFs) spanning 2,515 trading days, incorporating open-high-low-close (OHLC) prices, trading volume, and volume-weighted average price (VWAP) data. Simulations on assets such as NVDA, AAPL, and SPY successfully reproduce key empirical features, including heavy tailed return distributions, volatility clustering, and the absence of autocorrelation. For instance, Kolmogorov–Smirnov tests show that 98% of in-sample SPY simulations align with historical distributions at the 95% confidence level. Furthermore, out-of-sample evaluation over 277 trading days maintains 84% statistical consistency. These findings underscore the utility of regime-switching models in financial time series analysis, demonstrating how hybrid HMM based approaches can reveal latent market structures and sentiment dynamics under uncertainty.

2 - An Endogenized Controller Design for Speculative Market Dynamics

James Paine, Bucknell University Freeman College of Management, Lewisburg, PA, United States

Speculative bubble formation has been a long observed feature of markets, and continues to be a major focus of economic, marketing, and operations management research. While these bubbles may form because of specific physical or informational features of specific markets, it is also the interplay of current realized prices and future expected rewards, offset by rising risk of loss with increased speculation, that drives their formation. In other words, it can be the behavioral features of individual decision makers more so than the physics of these markets that yield these bubbles. The comparison of a current state of a system with some expected state in terms of instantaneous, forward projected, and cumulative observations is a feature of both this speculative bubble formation and classical control theory as well. This work presents a framework for adding more behaviorally realistic endogeneity into a classical proportional-integral-derivative (PID) controller design and illustrates its use in modeling speculative bubbles that both imbue and destroy value in an underlying commodity. More practically, this framework can be considered as the basis for a new fundamental unit modeling structure that bridges between control theory from operations research and speculative pricing theory from behavioral economics.

3 - Valuation, Weighted Average Cost of Capital, and Cost of Equity Under a Constant Leverage Ratio

Luis Chavez-Bedoya, ESAN Graduate School of Business, Lima, Peru

This study analyzes corporate valuation through the Weighted Average Cost of Capital (WACC) and Adjusted Present Value (APV) methods, assuming constant leverage. It thoroughly examines cost of equity and WACC calculation techniques, specifically investigating how constant (versus variable) discount rates affect investment valuations. The research identifies the cash flow conditions that validate using constant discount rates in financial modeling and develops a framework to properly discount future cash flows while isolating interest tax shield effects. Key contributions include: (1) methodological insights into maintaining fixed discount rates across valuation periods, (2) quantitative analysis of tax shield valuation under constant leverage, and (3) practical guidance for choosing between constant and variable rate approaches. The findings enhance understanding of capital structure effects on valuation accuracy in both WACC and APV contexts.

4 - CapOptix: An Options-Approach Based Framework for Capacity Market Pricing Models

Millend Roy, Columbia University, New York, NY, United States, Agostino Capponi, Vijay Modi, Vladimir Pyltsov, Yinbo Hu

Electricity and capacity markets increasingly face challenges in maintaining system reliability amid rising variability, investment uncertainty, and extreme demand-supply events. We present **CapOptix**, a pricing framework that redefines capacity market design through the lens of financial option theory. By modeling capacity commitments as **reliability options**—derivative contracts on wholesale electricity prices—CapOptix derives fair capacity premia under realistic stochastic dynamics, including Brownian motion, Geometric Brownian motion, GARCH models, jump-diffusion processes, and regime-switching frameworks. Departing from traditional mechanisms, CapOptix incorporates **Conditional Value-at-Risk (CVaR)** into its pricing architecture to capture rare but high-impact tail events, addressing risk dimensions that conventional expected-value metrics fail to reflect. The framework aligns incentives with system adequacy by rewarding capacity providers for mitigating extreme shortages rather than merely covering average conditions. A breakeven analysis shows that the derived premia closely approximate the **Net Cost of New Entry (NetCONE)**, while simultaneously identifying minimum contract durations necessary for investment viability.

CapOptix is empirically validated using real-world data from **Germany and NYISO**, two markets with distinct price dynamics. The approach demonstrates strong generalizability across European and U.S. contexts, supported by shortfall-driven price estimation calibrated via regression and evaluated through Monte Carlo simulation.

Overall, CapOptix delivers a robust, risk-sensitive, and flexible pricing mechanism for capacity markets. It offers practical tools for procurement under uncertainty and establishes a foundation for market reforms that can adapt to the evolving challenges of modern power systems.

5 - Greenium in Transition: Media Signals and the Pricing of ESG vs. Emission-Based Sustainability

Ruobing Yan, Harbin Institute of Technology, Harbin, China, People's Republic of, Qiang Ye

As awareness grows around the financial implications of climate change, sustainable investing is undergoing a strategic shift—from broad-based ESG integration to a more targeted focus on climate-specific metrics such as greenhouse gas (GHG) emissions. This research investigates how media signals—across both traditional news and social platforms—influence the pricing of sustainability in capital markets, specifically through the greenium: the differential in stock valuation between green and brown firms. Leveraging firm-level data, we compare greenium patterns when “greenness” is measured by subjective ESG scores versus objective GHG emission metrics. By integrating media sentiment and attention to “climate change” as explanatory variables, we model the informational pathways through which investor climate awareness moderates the relationship between sustainability indicators and market valuation, assessing how media-driven salience shapes investor responses to both perception-based and data-driven indicators of environmental performance. This study sheds light on the evolving dynamics of sustainable investment strategy, demonstrating how media systems condition market recognition of environmental risk and contribute to the strategic transition from ESG-based to climate-focused investing. The research bridges climate finance and information systems by translating sustainability indicators into financial valuations through media analytics, identifying the informational drivers of green asset pricing. Findings of the work would provide insights for investors seeking climate-hedged portfolios, regulators designing disclosure policies, and data providers developing the next generation of sustainability metrics in the post-ESG era.

ME05

Building A Level 3 A305

Well INFORMED PhD Retreat: Reclaiming Wellness in the Academy

Panel Session

Ambassadors Program

Co-Chair: Jessye Talley, Morgan State University, Baltimore, MD, United States

1 - Moderator Panelist

Jessye Talley, Morgan State University, Baltimore, MD, United States

Pursuing a doctorate is often marked by rigorous coursework, unrealistic expectations, and chronic stress—conditions that can push personal well-being to the sidelines. Unfortunately, this culture of self-sacrifice doesn't end at graduation. It often follows scholars into faculty life, where the pressure to publish, teach, and serve continues unchecked.

As part of our WELL INFORMED initiative, we're hosting a mini PhD retreat designed specifically for PhD students and early to mid-career faculty. This experience will include a restorative mix of discussion, reflective activities, and community connection, all aimed at helping you reimagine what wellness, productivity, and success can look like in academic life.

This isn't just about surviving the PhD or tenure track—it's about building a new, healthy, and effective normal within the academy

2 - Panelist

Orissa Massey, Sisters Scholars, Temple Hills, MD, United States

ME06

Building A Level 3 A306

Online Platform and Generative AI

Invited Session

eBusiness

Chair: Kan Xu, Arizona State University, Chandler, AZ, United States

1 - ChatGPT's Impact on Political Discourse: Evidence from Reddit

Xiaoning Wang, University of Texas at Dallas, Richardson, TX, United States

Large language models such as ChatGPT have transformed how people access information and create content. However, there has been little research on how these models affect individuals' political attitudes and expressions. In this paper, we investigate how the release of ChatGPT influenced political discourse on social media platforms. We collect data from the largest political discussion subreddit on Reddit and measure the political slant of each comment to infer each author's dynamic political stance. We find that the release of ChatGPT significantly intensified political polarization among users on the platform: users who were originally liberal-leaning became more liberal-leaning, and vice versa for those who were conservative-leaning. Further analysis reveals that this polarization was primarily driven by less active users who used ChatGPT to generate more extreme comments. Although active users did not appear to use ChatGPT themselves, they also became more active.

Interestingly, we find that ChatGPT did not reduce the level of civility on the platform or increase expressions of anger or disgust; instead, both hostility and toxicity on the platform declined, suggesting that while political polarization increased, affective polarization decreased, indicating that ChatGPT may have improved the quality of online political discourse.

2 - Does Platform Fragmentation Harm Innovation?

Yi Liu, University of Wisconsin-Madison, Madison, WI, United States, Zhuoxin Li

Platform fragmentation refers to the coexistence of multiple versions of a platform technology (e.g., Android 8 vs. Android 9) adopted by the users. The literature has highlighted the negative effect of fragmentation on innovation on the platform under the assumption of homogeneous developers, and found that fragmentation always harm innovation. In this paper, we consider heterogeneous developers, one big developer and one small developer. The big developer has cost advantage over the small one. Our analytical model shows that platform fragmentation is not always harmful. Platform fragmentation On one hand, platform fragmentation hinders complementary innovations by increasing developers' costs of serving multiple market fragments. Higher degree of platform fragmentation leads to less complementary innovations. On the other hand, platform fragmentation causes market segmentation and reduces the winner-take-all propensity. As a result, a higher degree of platform fragmentation is associated with more frequent innovations from small developers, and therefore, greater diversity of complementary innovations. An empirical study of the Firefox platform and its add-on market provides support of the analytical results.

3 - Learning User Engagement Using Deep Matching Embeddings in the Gaming Platform

Kan Xu, Arizona State University, Chandler, AZ, United States, Yu Qin, Olivia Sheng

In the competitive landscape of online gaming platforms, predicting short-term user engagement is critical for retaining users and maintaining platform success. Existing methods that rely on demographic data and gameplay statistics often fail to capture the complexity of user behaviors and interactions, limiting their ability to predict disengagement accurately. To address this, we propose a deep learning framework designed to model the gaming process and provide interpretable insights into why users may disengage. Our approach creates user embeddings by integrating diverse match-specific data, including game outcomes, in-game dialogue, and paralinguistic cues like emojis, to comprehensively capture user experiences. The model incorporates the game structure and user interactions, allowing it to explain potential causes for short-term user dropout. Using a multi-task learning strategy, the framework predicts engagement while considering behavioral and emotional signals, offering a nuanced view of factors influencing user retention. Crucially, our framework balances interpretability and predictive performance. It provides actionable insights that enable platforms to develop targeted retention strategies informed by a deeper understanding of user disengagement. Empirical evaluations using real-world gaming data show that our method significantly outperforms traditional benchmarks, demonstrating its effectiveness in predicting engagement and supporting policy decisions. This research

advances user engagement analytics and policy learning by introducing an interpretable, data-driven approach for predicting short-term engagement. The proposed framework serves as a decision-support tool that enables online gaming platforms to optimize business strategies, improve user retention, and drive revenue growth by proactively addressing the challenges of user disengagement.

4 - Emotional Dynamics: the Nonlinear Impact of Sentiment Volatility on Review Helpfulness

Jiaqi Ren, Zhejiang University, Hangzhou, China, People's Republic of, Xixian Peng, Yifan Yu, Yutong Guo

Affective cues embedded in online reviews play a pivotal role in shaping perceptions of review helpfulness. Prior research has primarily examined static affective characteristics, such as sentiment valence, arousal, and discrete emotions. In practice, emotional expressions often vary dynamically over time rather than remaining constant. This work introduces a novel affective cue: sentiment volatility, which captures the emotional dynamics in online reviews. Drawing on emotions as social information (EASI) theory, we explore how sentiment volatility influences perceived review helpfulness through both affective and inferential processes. We employ a multimethod approach involving (1) an analysis of a real-world dataset, (2) a functional near-infrared spectroscopy (fNIRS) experiment, and (3) an online experiment. Results of the three studies reveal an inverted U-shaped relationship between sentiment volatility and perceived review helpfulness, driven by two mechanisms: emotional contagion (affective process) and conceptual processing fluency (inferential process). In addition, we find that for the affective process, the negative effect of low sentiment volatility on emotional contagion is attenuated among readers with high levels of emotional contagion susceptibility. Our findings advance existing literature by highlighting emotional dynamics in the online review context and offer practical implications for platforms, reviewers, and merchants.

ME07

Building A Level 3 A307

Smooth and Convex Optimization

Job Market Showcase

Job Market Showcase

Chair: Alan Luner, Johns Hopkins University, Baltimore, 21210, United States

1 - Modern Convex Optimization & Hidden Convexity and Its Applications

Kevin Shu, California Institute of Technology, Pasadena, CA, United States

We often think about a dichotomy between convex and nonconvex optimization, the former being tractable and certifiably correct, and the latter being difficult but encompassing a much wider range of problems. This talk will present examples in which convex optimization can answer questions that do not on the surface seem to be convex, a surprisingly common phenomenon termed 'hidden convexity'. The two examples presented will be those of computer assisted design of algorithms and optimization over the manifold of rotation matrices, both of which arise in an array of applications and are far from being convex. First, we will describe how to construct first-order algorithms for optimization that have provably fast convergence rates using convex certificates of convergence, which are in some cases optimal with respect to strong metrics for convergence. Next, using ideas from topology, we show that optimization problems involving rotations of high-dimensional point clouds – i.e. optimization over the nonconvex set of rotation matrices -- can often be formulated as convex programs after a suitable low-dimensional projection; this is unlike the usual approach in which one 'lifts' a nonconvex problem to a convex one in higher dimensions. In both of these examples, the interaction with convex optimization facilitates mathematical guarantees for difficult optimization problems as well as fast algorithms for solving them.

2 - Spectral Methods for Polynomial Optimization

Elvira Moreno Ferreira, California Institute of Technology, Pasadena, CA, United States, Venkat Chandrasekaran

We propose a hierarchy of tractable relaxations based on spectral methods for polynomial optimization problems (POPs). Concretely, our hierarchy of spectral relaxations yields a monotonic sequence of bounds for the optimal value of a POP, each of which is computed as the minimum eigenvalue of a matrix obtained from the problem data. Because spectral methods are less computationally demanding than semidefinite programs, which underpin state-of-the-art techniques based on SOS certificates of non-negativity, our hierarchy provides an attractive alternative for obtaining bounds on large-scale problems. We identify the algebraic structure underlying a POP that makes it amenable to spectral relaxations, and we demonstrate the efficiency and applicability of our framework with numerical examples.

3 - A Performant Adaptive Linesearch-Free Subgame Perfect Gradient Method

Alan Luner, Johns Hopkins University, Baltimore, MD, United States

We present a highly performant gradient method for smooth convex optimization, drawing inspiration from several recent advances in the field. Our algorithm, the Adaptive Subgame Perfect Gradient Method (ASPGM) is based on the notion of subgame perfection, which provides a dynamic strengthening of the typical notion of minimax optimality. At each iteration, ASPGM makes a momentum-type update, optimized dynamically based on a (limited) memory/bundle of past function and gradient information. ASPGM is linesearch-free and fully adaptive due to its use of recently developed auto-conditioning, restarting, and preconditioning ideas. We show that ASPGM is performant on a wide range of problem classes and is competitive with BFGS and L-BFGS, the state of the art for these settings. Moreover, ASPGM maintains a certificate of solution quality, resulting in simple stopping criteria and restarting conditions.

ME08

Building A Level 3 A310

Technology and Society

Invited Session

Information Systems

Chair: Sameer Borwankar, McGill University, Westmount, QC, Canada

1 - Bridging Human Limitations: the Impact of Robot-Assisted Surgery on Clinical Outcomes and Healthcare Equity

Sung Joo Kim, Purdue University, West Lafayette, IN, United States

As AI is reshaping healthcare procedures and outcomes, Robot-assisted surgery (RAS) represents a cutting-edge advancement designed to transcend human limitations in both physical and cognitive aspects of surgical procedures, promising significant improvements in clinical outcomes. In this study, we investigate the impact of RAS on surgeon performance and healthcare outcomes. Especially, we explore how RAS implementation mitigates healthcare disparities as well as lowers in-hospital mortality rates.

2 - Building AI with Ai: the Impact of Generative Technologies on Startup Teams and Ai Product Innovation

Dongwook Chun, McGill University, Montreal, QC, Canada, Taha Havakhor, Animesh Animesh

This study investigates how the integration of artificial intelligence (AI) as a product feature—rather than merely as a development tool—shapes entrepreneurial team formation across diverse product domains. Leveraging data from over 76,000 early-stage product launches on ProductHunt (2021–2023), the paper introduces the concept of domain-level AI Functionality, measuring the functional alignment between product descriptions and AI patent technologies using contextualized topic modeling and semantic similarity. Employing a shift-share instrumental variable approach, the analysis reveals that domains with higher AI functionality traditionally required larger teams due to elevated demands for domain-specific knowledge depth and AI complexity. However, following the widespread access to Generative AI (GenAI) tools like ChatGPT, this relationship weakens, especially in generative, competitive, and enterprise-facing domains, as GenAI lowers the labor intensity and technical barriers of AI integration. While GenAI democratizes AI-enhanced product creation for smaller teams and solo entrepreneurs, it also reduces product quality differentiation and diminishes conceptual novelty. The findings illuminate how AI functionality and generative technologies are reshaping human capital demands and innovation patterns in entrepreneurial ecosystems, offering implications for founders, investors, and policymakers.

3 - Hyperparameter Optimization and Algorithm Selection: Practical Techniques, Theory, and New Frontiers

Dravyansh Sharma, Toyota Technological Institute, Chicago, IL, United States

Machine learning algorithms operate on data, and for any task the most effective method depends on the data at hand. Hyperparameter optimization and algorithm selection are therefore crucial to ensure the best performance in terms of accuracy, efficiency, reliability, interpretability, etc.

4 - Generative AI, Human Expertise, and Scaling Law

Jie Zheng, Purdue University, West Lafayette, IN, United States, Mohammad Rahman, Yu Hu

Evaluation is a critical component in business for producing high-quality knowledge outputs. It is typically conducted by human experts, but this process is difficult to scale, and expertise takes time to develop—making it scarce and often insufficient to assess the growing volume of knowledge outputs within organizations, especially with the broad use of generative AI. With the advancement of large language models (LLMs), there is growing interest in understanding whether LLMs can perform evaluations that are consistent with expert judgment. Because evaluation requires a high level of expertise—including both domain knowledge and expert judgment—we investigate how adding these types of data affects LLM evaluation. We partner with a consulting firm and design a two-stage experiment. In the first stage, human experts evaluate negotiation strategy reports. In the second stage, LLMs of different sizes are used to evaluate the same reports, with prompts adding either domain knowledge or expert judgment data. Our results show that expert judgment data improves agreement with experts more effectively than domain knowledge. We further demonstrate that expert judgment and domain knowledge impact agreement through within-role and across-role uncertainty in LLMs in different ways. In particular, higher levels of both types of uncertainty in LLMs can predict lower agreement with expert evaluations. Based on this insight, we propose a method that leverages uncertainty to select data, thereby improving agreement between LLMs and experts when scaling data.

ME09

Building A Level 3 A311

Novel AI Applications and Perspectives

Job Market Showcase

Job Market Showcase

Chair: Junyi Sha, MIT, Cambridge, MA, United States

1 - Human Decision-Making in AI Task Augmentation

Joseph Sebastian, Rensselaer Polytechnic Institute, Troy, NY, United States

In an era of rapid advancements in Artificial Intelligence (AI), effective human-AI collaboration is essential for enhancing decision-making, particularly when tasks involve varying levels of payout. While AI can perform increasingly complex tasks, determining when humans choose to delegate to AI remains an open question. This study investigates how humans delegate image classification tasks to AI in high- versus low-payout scenarios, and how trust in AI and individual risk preferences influence this behavior. Drawing on theories of absorptive capacity, algorithm aversion, and human-AI complementarities, we hypothesize that (1) individuals are more likely to delegate in low-payout scenarios, (2) risk attitude moderates the delegation decision, and (3) trust in AI further conditions this relationship. We conduct a two-part randomized experiment where participants classify images themselves or delegate to an AI agent, under varying financial incentives. Preliminary results reveal that delegation patterns vary significantly by payout structure, risk aversion, and trust levels, offering new insights into how augmented intelligence can be optimized in future workplace settings. These findings have theoretical and practical implications for the design of AI-integrated workflows and decision-support systems.

2 - Rethink AI for Vehicle Routing: Towards Generalization and Sustainability

Zhuoli Yin, Purdue University, West Lafayette, IN, United States

Artificial Intelligence (AI) promises to reshape combinatorial optimization problems like Traveling Salesman Problem and Vehicle Routing Problem. However, current AI based approaches face two key challenges: limited generalizability to real-world instances and the high environmental cost of training large models. In this talk, I present (1) a novel method leveraging vision language models to perform vision guidance for the optimization process without task-specific training, and (2) a new evaluation framework for quantifying the net sustainability benefits of applying AI to real-world vehicle routing.

3 - Productgpt: a Generative Model of Consumer Decision Dynamics in Limited-TIME Product Offerings

Jin Miao, The University of Texas at Dallas, Richardson, TX, United States, Fanglin Chen, Ying Xie

Managers frequently offer limited-time products across various industries, yet few studies examine consumers' dynamic decisions to guide managerial choices about reissuing previous products or introducing new ones. We develop *ProductGPT*, a novel deep generative pre-trained Transformer designed to predict sequential purchase decisions in limited-time product settings.

Three key features differentiate our *ProductGPT* from standard Transformer models. First, it uses triplet input tokens simultaneously encapsulate decision history, the supply-side sequence of limited-time products, and the demand-side sequence of observed decision outcomes. This approach enables *ProductGPT* to capture the rich and meaningful product representations for all products (i.e., both limited-time and always-available products) in the assortment. Second, using the encoded long-range and short-range contexts, *ProductGPT* exclusively predicts consumer decisions. During the inference phase, we can leverage the model to generate consumer decision sequences under alternative limited-time product schemes, thus providing managers a decision tool to help design limited-time products. Third, *ProductGPT* incorporates product features as covariates, enabling managers to forecast sales of new limited-time products through feature-based product embeddings.

We apply *ProductGPT* to a unique 20-month transaction dataset of video game loot boxes. Our model consistently outperforms several sophisticated benchmark models in sequential purchase predictions. We demonstrate its ability to forecast aggregate sales, discover behavioural patterns, and support product sequencing strategies.

4 - AI Tailoring: Evaluating Influence of Image Features on Fashion product Popularity

Junyi Sha, MIT, Cambridge, MA, United States

Identifying key product features that influence consumer preferences is essential in the fashion industry. In this study, we introduce a robust methodology to ascertain the most impactful features in fashion product images, utilizing past market sales data. First, we propose the metric called "influence score" to quantitatively assess the importance of product features. Then we develop a forecasting model, the Fashion Demand Predictor (FDP), which integrates Transformer-based models and Random Forest to predict market popularity based on product images. We employ image-editing diffusion models to modify these images and perform an ablation study, which validates the impact of the highest and lowest-scoring features on the model's popularity predictions. Additionally, we further validate these results through surveys that gather human rankings of preferences, confirming the accuracy of the FDP model's predictions and the efficacy of our method in identifying influential features. Notably, products enhanced with "good" features show marked improvements in predicted popularity over their modified counterparts. Our approach develops a fully automated and systematic framework for fashion image analysis that provides valuable guidance for downstream tasks such as fashion product design and marketing strategy development.

ME10

Building A Level 3 A312

Topics in Customer Choice

Job Market Showcase

Job Market Showcase

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

1 - Customer Preferences in Multi-Channel Services

Omer Berk Olmez, Baruch College, Graduate Center, CUNY, New York, NY, United States

As multi-channel services become more common, understanding the value customers place on virtual and in-person services has become increasingly important. Using data from a provider with in-person and virtual channels, we examine how service delivery channels and attributes—such as duration and distance—shape customer choices.

2 - Choice Modeling, Assortment Optimization, and Estimation when Customers are Non-Rational: Multinomial Logit Model with Non-Parametric Dominance

Yuheng Wang, Cornell University, New York, NY, United States, Dmitry Mitrofanov, Huseyin Topaloglu

Capturing non-rational choice behavior is crucial for accurately predicting customer choices and optimizing decision-making processes, as traditional choice models often do not capture non-rational factors that influence purchasing decisions. In this paper, we study a powerful two-stage choice modeling framework to capture a diverse array of non-rational customer behaviors. More specifically, we analyze a choice model where customers first use non-parametric dominance relationships to narrow down their options, and subsequently make a choice from the shortlisted products using the multinomial logit model. This choice model goes beyond the random utility maximization (RUM) framework by capturing a broad range of non-rational customer behaviors, including the decoy effect, choice overload, and preference cycles. Remarkably, we establish a clear and insightful connection between our choice model and the ranking-based framework that is equivalent to the RUM class. We demonstrate that our model can be represented using a ranking-based structure with negative 'probability' assigned to specific rankings. We then study the assortment optimization problem under this model, showing that it can be solved using a linear program. To fully operationalize this two-stage model, we also introduce its two-class version that captures the heterogeneity of customers in their propensity to deviate from the RUM framework. We demonstrate that the assortment optimization problem under this two-class variant is strongly NP-hard. Additionally, we propose an estimation framework to infer the dominance relationships in our model. Our numerical experiments highlight the model's superior predictive performance compared to RUM benchmarks and underscore its practical applicability in identifying revenue-maximizing assortments.

3 - Neural-Network Based Choice Model with Statistical and Optimality Guarantees

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

Discrete Choice Model has been a fundamental problem in operations, marketing, and economics to understand consumer choice behavior when provided a set of alternatives. The talk focuses on using neural networks for modeling rational choices and beyond, which tackles not only statistical properties and but also optimality guarantees. We demonstrate that overparameterization does not lead to overfitting when appropriate norm or entropy-regularization is applied. We prove that SGD-type algorithm achieves global convergence, despite the non-convexity of the problem. Empirical studies on datasets substantiate our theoretical findings.

ME11

Building A Level 3 A313

Machine Maintenance

Job Market Showcase

Job Market Showcase

Chair: Edikan Udofia, Colorado School of Mines, Lakewood, CO, United States

1 - Towards a Foundation Model for End-To-End Maintenance Optimization Using Sensor Data

MOHAMMAD BADFAR, Wayne State University, Detroit, MI, United States, Iman Kazemian, Murat Yildirim, Ratna Chinnam

Maintenance operations often dominate operational budgets—unplanned downtimes alone can cost millions per hour in equipment replacement and operational interruptions. Advances in sensor technologies enabled condition-based maintenance, where sensor-driven predictions on asset conditions inform optimal decision-making, yet the widespread deployment of such systems remains limited due to siloed, and expert-driven solutions that split into separate “predict” and “optimize” phases. We introduce ATOM, an end-to-end attention-based early-stage foundation model that learns directly from raw, multivariate sensor streams to prescribe optimal actions under real-world constraints, without ever needing to predict asset conditions. Built on a lightweight multi-head encoder-decoder, ATOM embeds resource limits, visit costs and crew availability via simple attention masks, collapsing forecasting and scheduling into a single policy network. On commercial Li-ion battery fleets, ATOM reduces total maintenance cost by 22–35% versus a state-of-the-art Predict-then-Optimize (PtO) pipeline, while closing over 90% of the gap to an Oracle with perfect foresight. We further show that by pretraining ATOM across randomly sampled cost structures, a single transferable agent can adapt to new cost regimes without retraining—demonstrating early-stages of a foundation-model behavior. Beyond batteries, our framework provides a four-step recipe for any domain that must convert streaming data into constrained decisions: (1) ingest raw features, (2) encode constraints via masks, (3) train on a task-aligned cost function, and (4) deploy a unified policy. Potential applications range from supply-chain replenishment to clinical resource scheduling and renewable-energy dispatch. By democratizing complex optimization behind an attention-powered interface, ATOM paves the way for foundation-model economics in operations across industries.

2 - Optimal Control of a Partially Observable Markov Deterioration Process

Cong Yang, University of British Columbia, Vancouver, BC, Canada, Hao Zhang

We study the optimal maintenance of a machine whose state is unobservable and deteriorates according to a discrete-time finite-state Markov process. At the beginning of any period, the machine may fail, which is self-announcing and will trigger an immediate corrective replacement. Otherwise, the decision maker can choose from three actions—production, inspection, and preventive replacement. Production is the normal action which continues the deterioration process; inspection reveals the machine state (deterioration level) perfectly; and replacement resets the machine to its best state. The model can be easily extended to incorporate more real-world features such as two types of failures and multiple replacement or repair options. It also describes an inventory management problem with unobserved inventory (or unrecorded demands). Adopting a relatively new methodology from the area of partially observable Markov decision processes, we identify several elegant structures of the optimal solution, which can be conveniently represented by graphs with multiple absorbing cycles. The optimal policies possess desired properties such as “dynamically scheduled inspections” and “postponed replacement after an inspection.” We develop a policy iteration algorithm to find the optimal solution exactly in finite time. This study fills a visible gap in the research of partially observable machine maintenance problems as well as partially observable inventory control problems.

3 - Approaches to Evaluate Condition Monitoring Systems for Risk Management in Manufacturing

Mehdi Dadfarnia, University of Maryland-College Park, College Park, MD, United States

Condition monitoring involves detecting, diagnosing, or predicting faults or failures in industrial equipment. Given advances in underlying AI solutions and IoT-based technologies, condition monitoring can rapidly improve maintenance processes. This talk discusses evaluation techniques and frameworks that justify investment in condition monitoring for manufacturing maintenance applications. This talk will present state-of-the-art evaluation methods for condition monitoring in industrial maintenance and discuss the challenges and current limitations of these methods. After presenting a procedural framework for evaluating the financial and engineering risks and benefits of condition monitoring systems in industrial applications, this talk introduces and compares simulation-based and analytical evaluation methods specific to monitoring for discrete-part manufacturing maintenance.

4 - Industrial Scale Data Analysis (Predictive Modeling of Force and Wear In Hot Rolling Continuous Steel Casting)

Edikan Udofia, Colorado School of Mines, Lakewood, CO, United States

This study develops a data analytics framework for predicting roll wear in hot steel rolling mills to optimize maintenance scheduling. We analyzed industrial datasets containing over 1GB of production data with 200+ variables per record to build predictive models for roll wear patterns.

ME12

Building A Level 3 A314

Job Market Candidates at the Research Frontier

Job Market Showcase

Job Market Showcase

Chair: Harriet Jeon, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States

1 - Incentivizing Smallholder Supplier Sustainability Under Behavioral Regularities**Yuan Shi, MIT, Boston, MA, United States, Iskandar Siregar, Yanchong Zheng**

Theoretical studies of sustainable farming incentives often assume rational decision-making, yet smallholder farmers exhibit behavioral biases that impact their decisions. Large-scale field experiments in this domain face measurement difficulties and are long term, high-stake efforts, which limits iterative testing of incentive structures. We address this by using *laboratory-in-the-field* experiments with Indonesian palm oil farmers, simulating real-world decision-making while maintaining experimental control. Results from our first experiment reveal key deviations from rational models: production-linked price premium incentives that are nearly optimal under the rational model did not outperform simple fixed payments due to behavioral factors including bounded rationality, cost aversion, and social preferences. To address this, we develop a behavioral incentive design framework, refining incentive structures and experimental design to align with observed decision-making patterns. In the redesigned second experiment, behaviorally calibrated premium contracts increased permanent sustainable adoption from 56% to 69% compared to fixed payments at no extra cost. Our findings highlight the need for empirically grounded, behaviorally informed incentives to enhance sustainability in supply chains and offer actionable insights to policymakers and practitioners.

2 - Adopting 'Green' Shareholder Proposals Can Improve Firm Operations**Hyungchan Cho, University of Michigan, Ann Arbor, MI, United States, Damian Beil, Andrew Wu**

This study examines whether the passage of environmental, social, and governance shareholder proposals leads to changes in firm operational performance. We show that firms whose environmental proposals pass experience statistically significant gains in operating efficiency: inventory turnover accelerates, cash conversion cycle shortens, and gross margin return on inventory rises, with effects emerging 3-5 years post-passage. In contrast, social and governance proposals display no consistent operational impact.

To understand the underlying channels, we first note that narrative disclosures from firms adopting environmental proposals often highlight commitments to invest in cleaner technologies, renewable power, and waste-reduction systems. Our analysis corroborates this, showing that treated firms significantly increase capital expenditures on property, plant, and equipment. Concurrent with these investments, we find a significant increase in asset retirements. This pattern suggests that the capital outlays are directed towards modernization and retrofitting rather than expansion. The resulting modernized asset base provides a plausible explanation for the observed delayed gains in operational performance.

The analysis leverages a proposal-level voting database for US public companies and applies a regression-discontinuity design; close wins and close losses supply quasi-random variation in adoption. We estimate dynamic treatment-on-the-treated effects by tracking firm outcomes from the meeting year through five subsequent fiscal years.

Our results suggest that the substance of a proposal matters: environmental measures translate into operational change, whereas social and governance measures do not. By linking shareholder activism to longer-run operating fundamentals, the paper provides a causal framework for understanding how specific ESG-driven mandates translate into firm-level operational performance.

3 - Rented Today, Bought Tomorrow: Buyout Pricing in the Circular Economy**Borja Apaolaza, The Wharton School - University of Pennsylvania, Philadelphia, PA, United States, Gerard Cachon, Santiago Gallino, Antonio Moreno**

Consumer preferences are shifting away from traditional ownership toward models that emphasize access, flexibility, and sustainability, fueling the growth of circular economy practices across industries. In the fashion sector, rental subscription services such as ThredUp, Nuuly, and Rent the Runway exemplify this transition by offering customers the ability to rent garments for a fixed monthly fee, with the added option to purchase rented items. This hybrid rental-sale model introduces a complex pricing challenge: setting buyout prices that balance immediate sales revenue with potential future rental income, while accounting for item-specific attributes such as popularity, condition, and customer price sensitivity.

In collaboration with a leading U.S. fashion rental company, we develop, estimate and validate a data-driven framework for optimizing buyout prices in this dual-revenue setting. We model the system using a Markov Decision Process (MDP) that captures the dynamics of garment rental, sale, and attrition, allowing us to prescribe revenue-maximizing prices while accounting for the complexities of the system. We calibrate our framework using real-world rental, purchase, and inventory data from our industry partner.

We validate our framework through extensive simulations and show that it significantly improves profitability compared to existing practices and alternative benchmarks. We estimate that the proposed pricing policy increases revenue by 3.1% over the company's current approach. Our analysis further highlights the importance of accounting for garment condition, and underscores the value of buyout options in managing inventory and generating additional income. Lastly, we propose a heuristic approach that delivers near-optimal performance with substantially reduced computational complexity.

ME13

Building A Level 3 A315

Allocation and Coordination

Job Market Showcase

Job Market Showcase

Chair: Jeff Sylvestre-Decary, University of Connecticut, Storrs, CT, 06084, United States

1 - Limits of Hegselmann-Krause Models**Aditya Gopalan, UIUC, Urbana, IL, United States, Partha Dey, Rasoul Etesami**

Hegselmann-Krause models are a localized, distributed averaging dynamics on spatial data. A key aspect of these dynamics is that they lead to cluster formation, which has important applications in geographic information systems, dynamic clustering algorithms, and social networks. For these models, the key questions are determining whether there exists a fixed point, and if so, characterizing it. In this talk, we will discuss two recent/ongoing works.

In the first, we discuss new results towards the “2R-Conjecture” for the Hegselmann-Krause model, for which no meaningful progress, or even any precise statement, has been made since its appearance in 2007. We provide, among other results, a proof that the expected cluster size converges to 2R and a statement of a stronger result about the convergence of the fixed points as the initial intensity tends to infinity, which is supported by simulation. In the second,

we discuss the existence of a fixed point of a stochastic variant of the Hegselmann-Krause dynamics which provides insights into algorithm development for large-scale dynamic clustering.

The key methodological contribution is to consider the dynamics as an infinite-dimensional problem on the space of point processes, rather than on finitely many points. This allows us to leverage stationarity and certain other symmetries to obtain the results. These techniques do not have finite-dimensional analogues.

2 - Designing Data Science Contests: the Role of Training Vs Test Splits

Ping-Chieh Huang, University of California-San Diego, La Jolla, CA, United States, Zhe Zhang, Sanjiv Erat

Companies organize data science contests to source innovative machine learning solutions for business operations. Using contests hosted on Kaggle as a motivating example, we formulate a model of data science contests where the organizer splits the data into a training part that is provided to contestants to build candidate models, and a testing part that is reserved privately to evaluate the submitted models. Previous literature in statistics suggests using intermediate data splits to strike a balance between having enough training data and being able to reliably rank models. However, by considering how participants choose their modeling approaches and the incentives they face in data science contests, we identify a new third effect of the split ratio. When intermediate splits are used — where the firm's evaluation of model performance is most reliable — contestants actually have the strongest incentive to submit “sub-optimal” models with higher variability in order to gain an advantage in the contest. An empirical examination of the past five years of Kaggle contests shows results that are consistent with our theoretical findings. Thus, an immediate implication of our model is that such intermediate split ratios advocated in prior statistics literature might result in the firm accurately choosing from a set of inaccurate submitted models.

3 - Practical Portfolio Optimization for Parlay Betting on Simultaneous Events

Jeff Sylvestre-Decary, University of Connecticut, Storrs, CT, United States, David Bergman, Bin Zou

This study addresses the challenge of optimal portfolio allocation in a classical betting setting, assuming the bettor seeks to maximize log-utility. While the Kelly Criterion is well-established as the optimal allocation for a single bet, questions remain open regarding how to proceed when considering multiple bets simultaneously. Building on prior literature, we present theoretical results on optimal parlay construction and propose a scalable logic-based Benders decomposition algorithm. Our formulation explicitly incorporates practical constraints faced by bettors—such as cardinality constraints on the number of bets placed and limits on parlay size—making the approach applicable in real-world sportsbook environments. We evaluate the algorithm's performance on both synthetic data and real-world sports betting instances, demonstrating its potential to enhance parlay strategies.

4 - Optimizing Mobility and Fairness in Public Housing Allocation

Yifeng Cao, Sauder School of Business, University of British Columbia, Vancouver, BC, Canada, Yichuan Ding, Daniel Granot, Hao Zhang, Weihua Zhang

We study a centralized public housing allocation problem, where a housing authority must assign limited housing units to applicants, including both move-in and transfer requests. Motivated by operational challenges faced by BC Housing, we propose a dynamic allocation framework that optimizes both mobility and fairness, where mobility is defined as the ease with which housing requests are fulfilled. Using and extending the matching queue framework of Gurvich and Ward (2014), we derive the optimal policy and adopt their policy for optimal review periods. In each period, the problem is formulated as a circulation problem with convex costs, solvable using a greedy algorithm. We prove that our algorithm is asymptotically optimal as arrival rates grow large, converging to an imbalance-based lower bound at an improved rate of $O(1/n^{1/3})$. Additionally, we determine the optimal priority to assign to transfer requests relative to move-in requests to achieve the optimal efficiency-fairness trade-off. Numerical experiments using real-world BC Housing data and simulations demonstrate that our algorithm achieves near-optimal performance, with solutions closely approximating the clairvoyant benchmark.

ME15

Building A Level 4 A401

Decision Making in Uncertain Dynamic Environments

Invited Session

Computing Society

Chair: Aras Selvi, Princeton University, Princeton, NJ, United States

1 - Breaking the Curse of Multiagency in Robust Multi-Agent Reinforcement Learning

Laixi Shi, California Institute of Technology (Caltech), Pasadena, CA, United States

Standard multi-agent reinforcement learning (MARL) algorithms are vulnerable to sim-to-real gaps. To address this, distributionally robust Markov games (RMGs) have been proposed to enhance robustness in MARL by optimizing the worst-case performance when game dynamics shift within a prescribed uncertainty set. RMGs remains under-explored, from reasonable problem formulation to the development of sample-efficient algorithms. Two notorious and open challenges are the formulation of the uncertainty set and whether the corresponding RMGs can overcome the curse of multiagency, where the sample complexity scales exponentially with the number of agents. In this work, we propose a natural class of RMGs inspired by behavioral economics, where each agent's uncertainty set is shaped by both the environment and the integrated behavior of other agents. We first establish the well-posedness of this class of RMGs by proving the existence of game-theoretic solutions such as robust Nash equilibria and coarse correlated equilibria (CCE). Assuming access to a generative model, we then introduce a sample-efficient algorithm for learning the CCE whose sample complexity scales polynomially with all relevant parameters. To the best of our knowledge, this is the first algorithm to break the curse of multiagency for RMGs, regardless of the uncertainty set formulation.

2 - Learning Causal Structure to Improve Sample Complexity in RL

Lily Xu, Columbia University, New York, NY, United States

One of the biggest barriers to real-world implementation of reinforcement learning (RL) algorithms is high sample complexity, especially when the action and state space are high in dimension. Here, we consider settings with multiple actions that each impact different subsets of state variables, which models many practical problems in healthcare, environmental management, and more. In many cases, the underlying structure is unknown a priori. We provide an approach to train a RL policy that simultaneously learns the underlying graph structure, enabling us to achieve strong performance within a shorter horizon.

3 - A Minimax Shrinkage Scheme for Wasserstein Distributionally Robust Optimization

Lingjun Meng, Imperial College Business School, London, United Kingdom, Ryan Cory-Wright, Wolfram Wiesemann

Wasserstein distributionally robust optimization (DRO) is a popular paradigm for decision-making under distributional ambiguity with a strong out-of-sample performance guarantee. While Wasserstein DRO sometimes admits tractable deterministic reformulations, it is still practically challenging to solve in many relevant applications. To enhance the scalability of Wasserstein DRO, this work proposes a scalable approximation algorithm for Wasserstein DRO. Namely, we interpolate between the optimal solution to a sample average approximation problem (i.e., a Wasserstein DRO problem with an ambiguity radius of zero) and a maximally robust optimization problem (i.e., a Wasserstein DRO problem with an ambiguity radius of infinity) to approximate the optimal Wasserstein DRO solution. We certify the quality of our solutions via a priori performance guarantee by providing an upper bound for the distance between the interpolated solution and the Wasserstein DRO solution. Additionally, we propose an instance-specific posteriori performance guarantee by interpolating both the primal and dual Wasserstein DRO problems and computing the gap between the primal and dual objective using the interpolated solution. Moreover, we develop a tractable bisection search scheme for selecting the optimal interpolation parameter. Finally, we demonstrate the practical efficacy of our approximation in several high-dimensional settings drawn from the finance and supply chain literature.

4 - Inventory Commingling on An E-Commerce Platform

Hanqi Wen, Ross School of Business, University of Michigan, Ann Arbor, MI, United States, Izak Duenyas, Joline Uichanco

Inventory commingling is an emerging e-commerce fulfillment strategy that is transforming how marketplace platforms with fulfillment services manage third-party seller inventories. This approach maintains identical products from multiple sellers in a shared virtual pool—where each unit is interchangeable regardless of physical location or original seller—thereby reducing the platform's shipping costs. While sellers benefit from the streamlined handling when commingling, they also face the risk that their customers may receive defective or substandard units from less scrupulous sellers exploiting the system.

We develop a two-stage game-theoretic model in which the platform first sets its fulfillment fee, and sellers with heterogeneous product qualities then decide whether to commingle, maintain separate inventories, or exit the platform. Our analysis reveals that lower-quality sellers are more likely to “free-ride” by commingling, introducing product quality risks that incentivize higher-quality sellers to keep separate inventories. Comparing equilibrium outcomes with and without commingling, we find that commingling consistently enhances platform profitability, although its impact on fulfillment fees and seller profits varies with market conditions. Implementing commingling in selective markets can ensure that commingling does not compromise product quality while realizing operational efficiencies. However, higher-quality sellers may suffer financially when the fulfillment fee rises to capture surplus gains.

5 - Going from a Representative Agent to Counterfactuals in Combinatorial Choice

Yanqiu Ruan, Singapore University of Technology and Design, Singapore, Singapore, Karthyek Murthy, Karthik Natarajan

We study decision-making problems where data comprises points from a collection of binary polytopes, capturing aggregate information stemming from various combinatorial selection environments. We propose a nonparametric approach for counterfactual inference in this setting based on a representative agent model, where the available data is viewed as arising from maximizing separable concave utility functions over the respective binary polytopes. Our first contribution is to precisely characterize the selection probabilities representable under this model and show that verifying the consistency of any given aggregated selection dataset reduces to solving a polynomial-sized linear program. Building on this characterization, we develop a nonparametric method for counterfactual prediction. When data is inconsistent with the model, finding a best-fitting approximation for prediction reduces to solving a compact mixed-integer convex program. Numerical experiments based on synthetic data demonstrate the method's flexibility, predictive accuracy, and strong representational power even under model misspecification.

ME16

Building A Level 4 A402

Forecasting and Optimization Methods and Models for Last Mile Operations

Invited Session

The Practice Section of Informs

Chair: Yaniv Mordecai, Amazon, Bellevue, WA, United States

1 - Exploring the AI Frontier in Optimization Research

Michael Wagner, University of Washington, Seattle, WA, United States, Rohit Malshe

We explore the ability of AI models to solve challenging optimization problems, focusing on evaluating their mathematical modeling skills rather than coding or writing abilities. The problems are taken from publicly available sources as well as originally written problems, from an introductory doctoral optimization course taught by Professor Michael Wagner at the University of Washington. The topics span various areas such as linear and integer programming, network optimization, dynamic programming, convex optimization, semidefinite programming, and robust optimization. We consider the most powerful AI models available in specific families: Anthropic's Claude 3 Opus, OpenAI's o1 pro, and Deepseek's R1. We ask the AIs to solve the problems, which are then assessed by Professor Wagner. We find that Opus solves 2/12 correctly, Deepseek-R1 solves 10/12 correctly, and o1 pro solves 11/12 correctly.

2 - Optimization Models Explained - Optix

Gokce Kahvecioglu, Amazon, Bellevue, WA, United States, Jin Ye, Ling Zhang

The opacity of decision-making processes in optimization models often hinders stakeholder trust and engagement. This paper introduces OptiX, a novel multi-agent large language model (LLM) framework to enhance the explainability of optimization models. OptiX addresses the critical gap between the technical capabilities of optimization models and their practical application, particularly in Amazon's vast supply chain ecosystem. OptiX is designed to provide clear, user-friendly narratives explaining optimization outputs to end-users with both technical and non-technical backgrounds. It highlights binding operational constraints, elucidates trade-offs guiding the model's recommendations, and assists scientists in rapid troubleshooting. The framework's architecture prioritizes simplicity and scalability, allowing it to adapt to various optimization models by consuming common artifacts such as solver logs, model formulations, and solutions, as well as science papers and technical documents. We assess the effectiveness of OptiX across a diverse set of optimization problems widely used at Amazon, including an high-impact optimization model in production where we allow the end-users interact directly with OptiX. Through systematic assessment by domain experts, our results demonstrate that OptiX generates explanations comparable in quality to those provided by human experts while delivering responses significantly faster.

3 - Assessing Vehicle Cleanliness Using Vision-Language Models

Girik Malik, Amazon, Austin, TX, United States, Jin Ye

As vehicle cleanliness becomes a critical aspect of safe vehicle operations, the demand for automated, objective solutions to assess and monitor cleanliness have become important. This research presents an innovative image-based scoring system designed to provide stakeholders with actionable insights into vehicle cleanliness. We demonstrate the efficacy of zero-shot learning in rating images based on their cleanliness levels. Furthermore, we demonstrate that our method can effectively align with human judgment based on only a limited number of in-context learning examples and produces results consistent with human scoring. By employing AI and image recognition techniques, the proposed system addresses the limitations of current methodologies, facilitating more efficient, safe, and hygienic vehicle operations. The system's potential extends beyond cleanliness monitoring, encompassing scalability for other applications such as vehicle damage detection. Future research will concentrate on fine-tuning VLM models to enhance human alignment with real-world scenarios and expanding the model's capabilities to operations beyond vehicle cleanliness through automated AI agents.

ME17

Building A Level 4 A403

Overcoming Challenges Moving from Data to Decisions

Invited Session

Practice Track

Chair: Somayeh Dejbord, Delta Air Lines, Atlanta, United States

1 - On TIME Every Flight: Saving Delays with Delta's Smart Gating

Ian Bailie-Sinal, Delta Air Lines, Atlanta, GA, United States

In the U.S. airlines' fiercely competitive landscape, operational performance is increasingly critical in meeting every customer's expectation of an on-time arrival. However, increases in flight volume and network complexity have made it difficult to synergize small operational decisions, stretching the human operator's ability to make globally optimal decisions in service of on-time performance. In this presentation, the Delta Air Lines (DAL) decision science team will detail how developing and implementing a flexible optimization model for gate assignments has helped address the scale and complexity of driving operational performance in Atlanta (ATL), the world's largest passenger airport. This talk will touch on both the technical challenges of building an at-scale optimization tool in DAL's largest hub, as well as the business challenges associated with developing a product that cuts across multiple operational divisions.

2 - Building the Foundation: Infrastructure for Data-Driven Decision at Scale

Arpit Rana, ACV Auctions, Buffalo, NY, United States

This talk will explore the challenges and lessons learned in designing a collaborative, scalable, and trusted data warehouse infrastructure to drive decision-making at ACV. Built on top of **Google BigQuery**, our OLAP-style architecture supported modular data models tailored for downstream reporting, forecasting, and machine learning initiatives.

We'll cover how the team implemented robust CI/CD practices using GitHub, Jenkins and Terraform to ensure consistency and maintainability, while orchestrating ELT workflows through Airflow to streamline and monitor data processing. BigQuery served as the analytical engine powering fast, reliable access to business-critical data.

Through real-world examples, we'll demonstrate how this modern data stack enabled analytics engineers, data scientists, and business stakeholders to identify cost inefficiencies, uncover operational insights, and drive product innovation through data-informed decisions.

3 - Practical Lessons from Evolving Lyft's Experimentation Platform

Anahita Hassanzadeh, Lyft, San Francisco, CA, United States

In this talk, we'll share practical lessons from building and scaling Lyft's experimentation platform and decision-making processes. We'll discuss challenges in scaling computational workflows, and the steps we've taken to move from data to decisions quickly, consistently, and with confidence.

4 - From Runway Configs to Taxi Time: a Seq2seq LSTM Approach at Delta Hubs

Zack Berman, Delta Air Lines, Atlanta, GA, United States, Benjamin Seavoy, James Grooms

Every day in the United States, over 45,000 flights take to the skies, with peaks exceeding 5,400 simultaneous aircraft. In this highly complex, safety-critical environment, even small gains in operational efficiency yield substantial cost savings and on-time performance improvements. To modernize core planning functions, Delta Air Lines has developed two complementary machine-learning models that forecast key surface-operations metrics hours in advance.

The first model predicts **Taxi-Out Time**, the interval between gate pushback and takeoff, using gradient-boosted decision trees trained on historical flight, airport, and weather data. Early taxi-out estimates enable optimized block-buffer allocation, more accurate fuel planning, and proactive recovery strategies under adverse conditions.

The second model forecasts **Runway Configuration** at major hub airports by leveraging a Long Short-Term Memory (LSTM) neural network. Ingesting high-dimensional meteorological and traffic-flow inputs, it accurately predicts which runway sets will be active several hours ahead, reducing surface-congestion risk and improving coordination between airline dispatch and air-traffic management.

Together, these applications illustrate how predictive analytics can drive measurable improvements in throughput and delay reduction. Attendees will learn model-development best practices, integration challenges in a regulated domain, and a roadmap for extending forecasting to other operational areas, forging a data-driven blueprint for next-generation airline planning.

ME18

Building A Level 4 A404

MAS Annual Awards

Award Session

Military and Security

Chair: Nathaniel Bastian, United States Military Academy, Chester, NY, United States

Co-Chair: Tahir Ekin, Texas State University, San Marcos, TX, United States

1 - Mas President's Introductory Remarks

Nathaniel Bastian, United States Military Academy, Chester, NY, United States

Welcome and Introduce awards

2 - Koopman Prize (Sponsored by MAS Members)- Presentation by Recipient

Rajan Batta, University at Buffalo, Buffalo, NY, United States

Koopman Prize (Sponsored by MAS Members)- Presentation by Recipient

3 - Seth Bonder Scholarship for Applied Operations Research in Military and Security Applications (Sponsored by the Bonder Estate) - Presentation by Recipient on Emerging Research Thread

Kyle Hunt, University at Buffalo, Lancaster, NY, United States

Seth Bonder Scholarship for Applied Operations Research in Military and Security Applications (Sponsored by the Bonder Estate) - Presentation by Recipient on Emerging Research Thread

4 - J. Steinhardt Prize (Sponsored by Cna Corporation) - Presentation by Recipient on Career Insights and Observations

Susan Sanchez, Naval Postgraduate School, Monterey, CA, United States

J. Steinhardt Prize (Sponsored by CNA Corporation) - Presentation by Recipient on Career Insights and Observations

ME19

Building A Level 4 A405

Markov Lecture

Award Session

Applied Probability Society

Chair: Neil Walton, University of Amsterdam, Amsterdam, Netherlands

1 - Going Beyond Mean-Field Approximations: Understanding High-Dimensional Stochastic Dynamics on Realistic Networks

Kavita Ramanan, Brown University, Providence, RI, United States

Large collections of randomly evolving particles that interact locally with respect to an underlying network model a variety of phenomena in engineering, operations research, machine learning and epidemiology. Classical work, falling under the rubric of mean-field approximations, has mostly focused on the case when this interaction graph is dense. However, most real-world networks are sparse and often random. We survey recently developed theory that provides insight into typical and atypical behavior in such systems.

2 - Discussant

Ruoyu Wu, Iowa State University, Ames, IA, United States

Discussant

ME20

Building A Level 4 A406

Round Table on Disruption Management for Transportation Companies

Invited Session

Railway Applications

Chair: Marc Meketon, Villanova University, Ambler, PA, 19002, United States

1 - Disruption Management at Delta Airlines

Ruhollah Heydari, Delta Air Lines, Atlanta, GA, United States

We will discuss how Delta Airlines handles disruptions: what is a disruption, what are our objectives in managing disruptions, our decision making/authority structures, technology that we use, and measuring our success at handling disruptions.

2 - Disruption Management at Norfolk Southern

Yannik Thomas, Norfolk Southern Corporation, Atlanta, GA, United States, Yudi Pranoto

We will discuss how Norfolk Southern handles disruptions: what is a disruption, what are our objectives in managing disruptions, our decision making/authority structures, technology that we use, and measuring our success at handling disruptions.

3 - Disruption Management at An Ltl Shipping Company

Jay Tomasello, Forward Air Corporation, Atlanta, GA, United States, Marc Meketon

We will discuss how a Less-than-truckload firm handles disruptions: what is a disruption, what are our objectives in managing disruptions, our decision making/authority structures, technology that we use, and measuring our success at handling disruptions.

4 - Disruption Management at UPS

Samaneh Shiri, UPS, Timonium, MD, United States

We will discuss how UPS handles disruptions: What is a disruption, what are our objectives in managing disruptions, our decision making/authority structures, technology that we use, and measuring our success at handling disruptions.

ME21

Building A Level 4 A407

Designing Smarter Urban Transit Systems

Contributed Session

Contributed

Chair: Tina Radvand, University of Illinois at Urbana- Champaign, Urbana, IL, United States

1 - Equitable Transportation Network Design for Signal-Free Smart Intersections

dan zhu, National University of Singapore, Singapore, Singapore, Tingting Xie, Yang Liu, Napat Rujeerapaiboon

Intersections often become bottlenecks, causing delays due to stop-and-go operations. Connected and autonomous vehicles (CAVs) promise smoother navigation through signal-free smart intersections. However, during the gradual transition from human-driven vehicles (HVs) to CAVs, we propose a hybrid intersection design (HID) that strategically integrates signal-free smart intersections with traditional signal-based ones by optimizing smart intersection locations and signal timings. Recognizing that HID may induce distributional welfare effects, especially disadvantaging HV users, we develop four bi-level programming models incorporating ethical principles—utilitarian, sufficient, difference, and maximax—to address equity concerns. In each model, transportation planners determine equitable HID decisions at the upper level, while travelers make user-optimal route choices at the lower level. We formulate the lower-level problem as a network equilibrium with embedded signal-free smart intersections under mixed traffic and derive its equivalent variational inequality (VI) formulation, proving VI solution existence. Notably, we show that under the difference principle, no traveler is worse off compared to fully signal-based control. To solve the models, we reformulate them as single-level mathematical programs with equilibrium constraints (MPECs) and approximate them by mixed-integer linear programs (MILPs). We further develop an efficient breakpoint selection technique and propose a modified augmented Lagrangian multiplier (MALM) approach to evaluate MILP solution quality. Numerical experiments validate the high effectiveness of the MILP solutions, achieving small optimality gaps (~1%), and show that HIDs significantly reduce travel times, particularly when CAV penetration reaches 80%. Moreover, signal-free smart intersections are preferably allocated to locations that experience severe signal and/or queue delays under signal-based intersection control.

2 - Optimal Micro-Transit Zoning Problem

Samitha Samaranayake, Cornell University, Ithaca, NY, United States, Hins Hu, Hongyi Jiang, Rhea Goswami

Micro-transit services offer a promising solution to enhance urban mobility and access, particularly by complementing existing public transit. However, effectively designing these services requires determining optimal service zones for the on-demand shuttles, a complex challenge often constrained by operational budgets and transit agency priorities. In our recent work, we propose two frameworks for designing optimal micro-transit service zones with an objective of maximizing served demand. The first framework is a novel two-phase algorithm. In Phase 1, it efficiently generates a comprehensive set of valid candidate zones, leveraging the notion of shareability graph, while in Phase 2, it formulates the selection of the optimal zones as a Budget-Constrained Weighted Maximum Coverage Problem, which is easily solvable via integer programming. The second framework adopts a column generation approach, providing exact solutions and allowing integration of pricing heuristics for improved efficiency. We evaluate both frameworks on synthetic and real-world datasets, validating their effectiveness and scalability and demonstrating their superior performance compared to a baseline algorithm.

3 - A Goal Programming Approach for Optimizing Microtransit in Rural Areas

Tahmina Sultana, Clemson University, Clemson, SC, United States, Pamela Murray-Tuite

Limited access to dependable and adaptable transit remains a major barrier in rural areas across the United States. We propose a shared microtransit optimization model developed to improve mobility where traditional fixed-route systems are insufficient. Designed from both user and provider perspectives, the model balances passenger needs with operational efficiency. The model has three core goals: minimizing unserved ride requests, passenger travel time, and overall operational costs. It operates over discrete time steps. Passenger assignments are made based on origin-destination needs, time windows, and vehicle capacity. This model offers a practical tool for transit and demand response service planners and researchers aiming to develop flexible transportation solutions in rural regions.

4 - A Modeling Framework to Generate Activity-Based Travel Demand from Trip-Based Model

Yueshuai He, University of Louisville, Louisville, KY, United States, Eronmonsele Esekhaigbe, Lihui Bai

Activity-based models (ABMs) are widely recognized as the state-of-the-art approach for travel demand forecasting, offering detailed representations of individual behavior through interdependent sequences of activities. While ABMs offer clear advantages over traditional trip-based models—especially for evaluating complex policies and long-range planning—they remain underutilized due to extensive data and computational demands. This study presents a modeling framework for generating activity-based travel demand by leveraging outputs from an existing trip-based model. The objective is to support the broader adoption of ABMs while maintaining analytical rigor.

A case study is conducted for the Kentuckiana Regional Planning & Development Agency (KIPDA). Using the 2000 KIPDA household travel survey as a behavioral foundation, the framework begins with the generation of a synthetic population for the target year using iterative proportional fitting algorithm, aligning American Community Survey data with key socioeconomic and travel-related characteristics. Next, unique activity chains are derived at the individual level from the 2000 KIPDA household travel survey and linked to demographic profiles. These chains are expanded across the calibrated synthetic population to replicate representative travel patterns. Finally, a maximum likelihood optimization model is developed to assign spatial locations to each activity in a way that matches the origin-destination patterns from the trip-based model.

By reusing legacy survey data and existing model infrastructure, this framework offers a cost-effective, scalable pathway for researchers to transition toward more behaviorally realistic and policy-sensitive demand models across diverse geographic regions.

5 - Optimization of a Feeder Transit System Under Different Market Regimes

Jiaqing Lu, Florida State University, Tallahassee, FL, United States, Qianwen Guo

This research addresses the critical need for optimizing feeder transit systems that connect expanding suburban areas to major transit hubs, a challenge exacerbated by increasing urban population and traffic congestion. The study focuses on determining optimal values for key planning parameters, including feeder line density within a linear transportation corridor, stop spacing along routes, fare structures, and service headway. These parameters are optimized under distinct market regimes, specifically operator profit maximization and social welfare maximization. Two analytical optimization models are developed, incorporating an elastic travel demand function sensitive to service attributes. These models rigorously explore the intricate relationships between suburban population distribution, the design of transit lines and stop configurations, and overall system performance. A significant contribution of this work is the explicit consideration of different fare schemes (zone-based and distance-based fares) within the optimization framework. The research systematically investigates how varying population densities influence optimal line and stop spacing strategies (e.g., uniform versus uneven spacing) and how these parameters, in turn, affect operator profit and social welfare. Furthermore, the study will apply these proposed maximization models to two real-world feeder transit systems: the services in Denton County and Miami-Dade County. This application aims to determine optimal operational parameters tailored to the specific characteristics and diverse scales of these transit environments, thereby demonstrating the models' practical applicability for enhancing feeder service efficiency and effectiveness.

6 - How Fare Capping Affects Transit Ridership and Social Welfare? An Optimization Approach with Elastic and Spatially Distributed Demand

Ziyue Li, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Qianwen Guo

Fare capping, where riders are not charged more than a specified maximum over a day, week, or month regardless of the number of trips, has been successfully adopted by several major transit agencies. Empirical evidence from agencies such as Miami-Dade Transit indicates that fare capping can help attract additional ridership. However, theoretical understanding of its broader impacts remains limited. Specifically, there is a lack of microeconomic models that capture how fare capping influences both ridership behavior and agency revenue. To address this gap, this paper focuses on distance-based fare capping, where fares increase with distance traveled but are subject to a maximum cumulative limit. An optimization model is proposed to design fare capping policies that maximize social welfare. The model assumes that transit agencies set fare structures while riders respond with elastic demand to fare changes. Optimal fare capping policies are derived, and their effects on ridership, revenue, and social welfare are analyzed. Numerical experiments using data from Miami-Dade Transit offer practical insights for agencies seeking to implement equitable and effective fare capping strategies.

7 - Rebalancing Modular Transit Systems: Graph-Based Optimization for Fleet Size Planning and Routing

Tina Radvand, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Alireza Talebpour, Yanfeng Ouyang

Modular bus systems aim to address the scalability and flexibility challenges of current transit systems. Rebalancing the distribution of empty pods in modular bus systems is crucial for minimizing operational costs and improving system efficiency. This study aims to offer a solution to the rebalancing problem by reducing this problem to a minimum path cover and minimum cost circulation problems. The minimum fleet size and optimal pod routings are determined in the proposed approach. The unique structure of modular bus systems allows these optimization problems to be solved efficiently in polynomial time.

ME22

Building A Level 4 A408

Theoretical Foundations of Reinforcement Learning

Invited Session

Applied Probability Society

Chair: Sajad Khodadadian, Virginia Tech, Blacksburg, VA, 24061, United States

1 - Contextual Online Pricing with (Biased) Offline Data

Yixuan Zhang, University of Wisconsin-Madison, Madison, WI, United States, Ruihao Zhu, Qiaomin Xie

We study contextual online pricing with biased offline data. For the scalar price elasticity case, we identify the instance-dependent quantity Δ^2 that measures how far the offline data lies from the (unknown) online optimum. We show that the time length T , bias bound V , size N and dispersion $\lambda_{\min}(\hat{\Sigma})$ of the offline data, and Δ^2 jointly determine the statistical complexity. An Optimism-in-the-Face-of-Uncertainty (OFU) policy achieves a minimax-optimal, instance-dependent regret bound $\tilde{O}(\sqrt{T} \wedge (V^2T + \frac{dT}{\lambda_{\min}(\hat{\Sigma})}) + (N \wedge T) \Delta^2)$. For general price elasticity, we establish a worst-case, minimax-optimal rate $\tilde{O}(\sqrt{T} \wedge (V^2T + \frac{dT}{\lambda_{\min}(\hat{\Sigma})}))$ and provide a generalized OFU algorithm that attains it. When the bias bound V is unknown, we design a robust variant that always guarantees sub-linear regret and strictly improves on purely online methods whenever the exact bias is small. These results deliver the first tight regret guarantees for contextual pricing in the presence of biased offline data. Our techniques also transfer verbatim to stochastic linear bandits with biased offline data, yielding analogous bounds.

2 - Sample Complexity of Offline Distributionally Robust Linear Markov Decision Processes

He Wang, Carnegie Mellon University, Pittsburgh, PA, United States, Yuejie Chi

In offline reinforcement learning (RL), the absence of active exploration necessitates robust modeling to address the sim-to-real gap, where discrepancies between the simulated and deployed environments can significantly degrade the performance of the learned policy. To equip the policy with robustness in a sample-efficient manner—especially in high-dimensional state-action spaces—this paper investigates the sample complexity of distributionally robust linear Markov decision processes (MDPs), using offline data and an uncertainty set defined by total variation distance. We develop a pessimistic model-based algorithm and establish its sample complexity bound under minimal data coverage assumptions, improving upon prior work by at least a factor of $O(d)$, where d is the feature dimension. We further enhance the performance guarantee of the proposed algorithm by incorporating a carefully designed variance estimator.

3 - Pilaf: Optimal Human Preference Sampling for Reward Modeling

Yaqi Duan, New York University, Stern School of Business, New York, NY, United States

As large language models increasingly drive real-world applications, aligning them with human values becomes paramount. Reinforcement Learning from Human Feedback (RLHF) has emerged as a key technique, translating preference data into reward models when oracle human values remain inaccessible. In practice, RLHF mostly relies on approximate reward models, which may not consistently guide the policy toward maximizing the underlying human values. We propose Policy-Interpolated Learning for Aligned Feedback (PILAF), a novel response sampling strategy for preference labeling that explicitly aligns preference learning with maximizing the underlying oracle reward. PILAF is theoretically grounded, demonstrating optimality from both an optimization and a statistical perspective. The method is straightforward to implement and demonstrates strong performance in iterative and online RLHF settings where feedback curation is critical.

4 - Online Learning in Risk Sensitive Constrained Mdp

Mehrdad Moharrami, University of Iowa, Iowa City, IA, United States, Arnob Ghosh

We consider a setting in which the agent aims to maximize the expected cumulative reward, subject to a constraint that the entropic risk of the total utility exceeds a given threshold. Unlike the risk-neutral case, standard primal-dual approaches fail to directly yield regret and violation bounds, as value iteration with respect to a combined state-action value function is not applicable in the risk-sensitive setting. To address this, we adopt the Optimized Certainty Equivalent (OCE) representation of the entropic risk measure and reformulate the problem by augmenting the state space with a continuous budget variable. We then propose a primal-dual algorithm tailored to this augmented formulation. In contrast to the standard approach for risk-neutral CMDPs, our method incorporates a truncated dual update to account for the possible absence of strong duality. We show that the proposed algorithm achieves regret of $\tilde{O}(V_{g,\max} \cdot K^{\frac{3}{4}} + \sqrt{(H^4 \cdot S^2 \cdot A \cdot \log(1/\delta)) \cdot K^{\frac{3}{4}}})$ and constraint violation of $\tilde{O}(V_{g,\max} \cdot \sqrt{(H^4 \cdot S^2 \cdot A \cdot \log(1/\delta)) \cdot K^{\frac{3}{4}}})$ with probability at least $1 - \delta$, where S and A denote the cardinalities of the state and action spaces, respectively, H is the episode length, K is the number of episodes, $\alpha < 0$ is the risk-aversion parameter, and $V_{g,\max} = (1/\alpha)(\exp(\alpha H) - 1)$. To the best of our knowledge, this is the first result establishing sublinear regret and violation bounds for the risk-sensitive CMDP problem.

ME24

Building A Level 4 A410

Data-Driven Decision Making for Social Impact

Invited Session

Applied Probability Society

Chair: Stefan Wager, Stanford GSB, Stanford, CA, United States

1 - Lfms for Education

Hamsa Bastani, Wharton School, Philadelphia, PA, United States, Tsai-Hsuan (Angel) Chung, Osbert Bastani

We will discuss recent results from large-scale randomized controlled trials leveraging large language models for education. We show how we can improve students' long-term learning outcomes by incorporating guardrails and intentional optimization of system prompts and capabilities.

2 - The Cost of Ending Extreme Poverty

Roshni Sahoo, Stanford University, Stanford, CA, United States

What would it cost to end extreme poverty? To obtain realistic and actionable cost estimates, we study poverty targeting as a statistical learning problem. In contrast to existing estimates of the global poverty gap, which assume that policymakers can perfectly target the minimum transfer required to lift each household above the poverty line, our estimates assume that policymakers have imperfect information and imperfect ability to target. In a sample of 13 countries that collectively account for 35% of the world's poor, the cost of reducing the population-weighted poverty gap index to 1% via a cash-transfer policy that targets based on observable characteristics is \$67.6B nominal per year. This cost is 2.78 times that of an oracle policy that perfectly targets the minimum transfer required to lift households above the poverty line but is 28% of the cost of universal basic income at the level of the poverty line.

3 - Markovian Search with Ex-Ante Constraints: Theory and Applications in Socially-Aware Algorithmic Hiring

Rad Niazadeh, Chicago Booth School of Business, Chicago, IL, United States

Re-designing and calibrating decision-making processes to achieve goals of equity and access in high-stakes applications, such as hiring, is one of the pillars of algorithmic fairness. Most of the literature focuses on machine learning problems in these contexts, and less attention has been given to calibrating operational decisions with operational intricacies, e.g., sequential search and selection decisions in the presences of resource limitations.

In this talk, motivated by socially-aware applications in algorithmic hiring (e.g., making the search processes fair and diverse), I will elaborate on an algorithmic framework to incorporate ex-ante constraints into "stateful" sequential search problems. Our framework includes the classic Weitzman's Pandora's box problem [Weitzman, 1978] and its extensions to joint Markovian scheduling [Dumitriu et al., 2003; Gittins, 1979], which handle richer search processes (e.g., multistage search with rejection). We show how the optimal policy in the presence of a single ex-ante constraint (such as demographic parity) retains an index-based structure similar to the unconstrained case, but may randomized between two dual-based adjustments that are both easy to compute and economically interpretable. We then extend our results to handle multiple affine constraints by reducing the problem to a variant of the exact Carathéodory problem and providing a novel polynomial-time algorithm to generate an optimal randomized dual-adjusted index-based policy that satisfies all constraints simultaneously. Next, by imposing general affine and convex ex-ante constraints, we develop a primal-dual algorithm that randomizes over a polynomial number of dual-based adjustments to the unconstrained indices, yielding a near-feasible, near-optimal policy.

4 - Value-Based Learning with Misspecified Reward Models

Peter Frazier, Cornell University, Ithaca, NY, United States, Samuel Tan

In offline contextual bandits, value-based learning trains a reward model as a function of context and executes the greedy policy with respect to the predicted reward. We study the loss function used to train the reward model and its impact on regret. Mean squared error (MSE), the typical choice, optimizes for predictive accuracy and not regret. Does this matter? A recent paper, Hu et al. 2024, seems to suggest that the answer is no, showing that the regret of value-based learning converges to the best regret achievable in the induced policy class, even when the reward model class is misspecified. But we argue that the answer is yes. By violating a critical assumption in Hu et al. 2024, we give an example of a misspecified reward model class in which the regret of value-based learning trained with MSE fails to converge to that of the best model in the reward model class. To address this limitation of MSE, we present a novel loss function, Empirical Soft Regret (ESR), for value-based learning that specifically targets regret rather than accuracy. When actions are binary, we prove that ESR asymptotically matches the best-in-class regret even when the reward model is misspecified. ESR is differentiable, supporting flexible machine learning model classes dependent on gradient-based training. In experiments on real-world decision-making problems from news recommendation and personalized healthcare, ESR significantly outperforms state-of-the-art algorithms from contextual bandits and conditional average treatment effect estimation.

5 - When Survival Meets Causal: A Framework for Decision Making on Time-to-Event Outcomes

Jessy Xinyi Han, Massachusetts Institute of Technology, Cambridge, MA, United States

Understanding not just *whether* an event occurs but *when* it happens—and how its occurrence hazard evolves over time—is central to many fields, from patient survival in healthcare and equipment failure in engineering to criminal recidivism in justice systems and customer churn in insurance. Classical survival analysis provides valuable tools for this task but often conflates association with causation and suffers from confounding bias in observational settings where treatment assignment is non-random. As a result, it cannot reliably predict outcomes under alternative interventions or policies, limiting its utility for informed decision-making.

This talk first introduces our methodological contribution, Synthetic Survival Control (SSC), a novel tensor-based approach for causal survival analysis that estimates counterfactual hazard trajectories by synthesizing weighted survival paths from comparable units. SSC enables identification in complex, real-world panel data with heterogeneous treatments and confounding, with finite-sample guarantees.

Next, I demonstrate the application of SSC and causal survival analysis through two empirical examples. In healthcare, analysis of T-cell blood cancer patient data with staggered therapy rollouts reveals clear survival improvements from new treatments. In criminal justice, we show that racial disparities in recidivism are not fully explained by risk assessment tools alone—broader contextual factors, including socioeconomic ones like unequal access to support systems, also play a critical role, underscoring important avenues for policy intervention.

By bridging survival analysis and causal inference, this framework offers interpretable, data-driven insights with broad policy implications for improving outcomes across healthcare, justice, and policy.

ME27

Building B Level 2 B201

GPU Acceleration of Optimization Algorithms I

Invited Session

Computing Society

Chair: Matthew Galati, Amazon, Glen Mills, PA, United States

1 - Opportunities and Challenges for Lp and Mip on the Gpu**Robert Luce, Gurobi Optimization, Beaverton, OR, United States**

Recent GPU hardware generations offer high compute throughput for many applications in machine learning, scientific computing, and beyond. How does this new compute power translate into improvements in solving real-world linear and mixed-integer optimization problems in practice?

We review Gurobi's efforts and research findings on solving optimization problems on the GPU. We touch interior point methods, first-order methods, and relaxation-based MIP heuristics, and we frame these in terms of their applicability and usability on the GPU. Finally, we'll showcase how Gurobi's preliminary GPU-based components compare, performance-wise, to our standard CPU-based counterparts across a wide range of real-world models

2 - Gpu Acceleration in the Cardinal Optimizer**Nils-Christian Kempke, COPT GmbH, Berlin, Germany**

The Cardinal Optimizer (COPT) is a high-performance solver designed for large-scale optimization problems, with support for linear programming (LP), mixed-integer programming (MIP), and conic programming. In this talk, we present recent advancements in COPT's GPU-accelerated capabilities, discuss key computational and theoretical enhancements, and share performance benchmarks from the latest release.

3 - Pdlp: a Practical First-Order Method for Large-Scale Linear Programming**Miles Lubin, Hudson River Trading, New York, NY, United States, David Applegate, Mateo Diaz, Oliver Hinder, Haihao Lu, Brendan O'Donoghue, Warren Schudy**

We present PDLP, a practical first-order method for linear programming (LP) designed to solve large-scale LP problems. PDLP is based on the primal-dual hybrid gradient (PDHG) method applied to the minimax formulation of LP. The talk focuses on our more recent contributions since the 2021 NeurIPS paper, namely: the C++ implementation in Google's OR-Tools, the feasibility polishing heuristic, and a new collection of large-scale test instances.

4 - Parallel by Design : Gpu-Powered Optimization at Nvidia**Chris Maes, Nvidia, San Francisco, CA, United States**

This talk gives an overview of cuOpt, NVIDIA's GPU-accelerated solver for linear (LP) and mixed-integer (MIP) programs. We focus on two GPU-native LP solvers: a parallel implementation of Mehrotra's predictor–corrector method and PDLP, a recent first-order method based on the primal–dual hybrid gradient algorithm. We also highlight advances in GPU-based heuristics for MIP and discuss how GPU and CPU algorithms can be combined to solve a wide range of optimization problems.

5 - GPU-Accelerated Optimization Approaches with FICO(R) Xpress**Dinakar Gade, FICO, Bozeman, MT, United States, Michael Perregaard**

This talk will present observations on the integration of GPU-acceleration techniques using FICO(R) Xpress Solver. We will present early-stage insights into the use of GPU-based algorithms with a focus on the primal-dual hybrid-gradient algorithms for LP solving and areas where GPUs acceleration can be exploited. We will present considerations for hardware and software development and discuss future avenues to improve upon these methods.

ME28

Building B Level 2 B202

Conic, Combinatorial and Global Optimization

Invited Session

Computing Society

Chair: Boshi Yang, Clemson University, Clemson, SC, United States

Co-Chair: Hao Hu, Clemson University, Clemson, SC, United States

1 - Facial Reduction in Biqcrunch**Ian Sugrue, Morehead State University , Morehead, KY, United States**

BiqCrunch is a binary quadratic solver that uses a semidefinite bounding procedure with a branch-and-bound framework to get exact solutions to combinatorial optimization problems. Since its latest official update, BiqCrunch has undergone further development, most notably the inclusion of the dimensionality reduction process known as facial reduction. We will discuss the details of how facial reduction is implemented, the numerical results that support this inclusion, as well as future development plans for BiqCrunch.

2 - Facial Reduction for Semidefinite Relaxations of Combinatorial Optimization Problems**Hao Hu, Clemson University, Clemson, SC, United States, Mingming Xu, Boshi Yang**

In this talk, we present new findings on facial reduction for semidefinite relaxations of combinatorial optimization problems. In semidefinite programming (SDP), Slater's condition is crucial for both theoretical convergence guarantees and the practical performance of optimization algorithms. When Slater's condition fails, facial reduction can restore it through a finite sequence of reformulations. However, these reformulations often involve solving auxiliary optimization problems that can be as challenging as the original. Recent research has therefore focused on developing more efficient strategies for performing facial reduction.

In our work, we specifically consider SDP problems that arise as relaxations of combinatorial optimization problems. This perspective enables us to exploit the underlying combinatorial structure, allowing the development of novel and highly efficient facial reduction techniques. We establish theoretical results demonstrating the effectiveness of our approach. Numerical experiments further show that applying our specialized facial reduction method significantly improves both the speed and accuracy of solving SDP problems.

3 - SDP Approach to Quadratic K-Vertex-Disjoint Paths Problem**Mingming Xu, Clemson University, Clemson, SC, United States, Hao Hu**

We consider the quadratic k -vertex–disjoint paths problem (Q-VDP). Given a directed graph G and k source–target pairs $(s_1, t_1) \dots (s_k, t_k)$, the objective is to find k mutually vertex–disjoint paths that minimize a quadratic cost function. Such quadratic objectives naturally arise when interactions among paths matter. The problem is NP-hard when k is part of the input. In this work, we formulate Q-VDP as an integer quadratic program and derive semidefinite-programming (SDP) relaxations. We also develop a

tailored Alternating Direction Method of Multipliers (ADMM) to solve these relaxations, yielding a fast algorithm that recovers exact solutions in practice. Numerical results show that our method finds optimal solutions faster than the commercial solver Gurobi.

4 - Cuhallar: a Gpu Accelerated Low-Rank Augmented Lagrangian Method for Large-Scale Semidefinite Programming

Arnesh Sujarani, University of Waterloo, Waterloo, ON, Canada, Renato Monteiro

This talk discusses cuHALLaR, a GPU-accelerated implementation of a solver for solving large-scale semidefinite programming (SDP) problems. We demonstrate how our Julia-based implementation efficiently uses GPU parallelism through optimization of simple, but key, operations, including linear maps, adjoints, and gradient evaluations. Extensive numerical experiments across three problem classes—maximum stable set, matrix completion, and phase retrieval show significant performance improvements over both CPU implementations and existing GPU-based solvers. For the largest instances, cuHALLaR achieves speedups of 30-140x on matrix completion problems, up to 135x on maximum stable set problems for Hamming graphs with 8.4 million vertices, and 15-47x on phase retrieval problems with dimensions up to 3.2 million. cuHALLaR efficiently handles massive problems with matrix variable of dimension up to 8 million and number of constraints up to 300 million at 10^{-5} relative precision. It matrix completion instances with over 8 million rows and columns in just 142 seconds. These results establish cuHALLaR as a very promising GPU-based method for solving large-scale SDPs.

5 - Convexification of Bilinearly Constrained Set

Hyun-Ju Oh, Clemson University, Clemson, SC, United States, Margaret Wiecek, Boshi Yang

In this talk, we present the convexification of sets defined by box constraints and bilinear inequalities, in which all bilinear monomials share a common variable. We show the polyhedrality of the convex hull of such sets and present its linear and vertex descriptions.

ME29

Building B Level 2 B203

ENRE Awards Session

Award Session

Energy, Natural Resources and the Environment

Chair: Benjamin Leibowicz, The University of Texas at Austin, Austin, TX, United States

Co-Chair: Gokce Palak, Wake Forest University, Winston-Salem, NC, United States

1 - Enre Student Best Paper Award: Adaptive Robust Optimization with Data-Driven Uncertainty for Enhancing Distribution System Resilience

Shuyi Chen, Carnegie Mellon University, Pittsburgh, PA, United States

Extreme weather events are placing growing strain on electric power systems, exposing the limitations of purely reactive responses and prompting the need for proactive resilience planning. However, existing approaches often rely on simplified uncertainty models and decouple proactive and reactive decisions, overlooking their critical interdependence. This paper proposes a novel tri-level optimization framework that integrates proactive infrastructure investment, adversarial modeling of spatio-temporal disruptions, and adaptive reactive response. We construct high-probability, distribution-free uncertainty sets using conformal prediction to capture complex and data-scarce outage patterns. To solve the resulting nested decision problem, we derive a bi-level reformulation via strong duality and develop a scalable Benders decomposition algorithm. Experiments on both real and synthetic data demonstrate that our approach consistently outperforms conventional robust and two-stage methods, achieving lower worst-case losses and more efficient resource allocation, especially under tight operational constraints and large-scale uncertainty.

2 - Enre Early Career Best Paper Award: Regression Equilibrium in Electricity Markets

Vladimir Dvorkin, University of Michigan, Ann Arbor, MI, United States

In two-stage electricity markets, renewable power producers enter the day-ahead market with a forecast of future power generation and then reconcile any forecast deviation in the real-time market at a penalty. The choice of the forecast model is thus an important strategy decision for renewable power producers as it affects financial performance. In electricity markets with large shares of renewable generation, the choice of the forecast model impacts not only individual performance but also outcomes for other producers. In this paper, we argue for the existence of a competitive regression equilibrium in two-stage electricity markets in terms of the parameters of private forecast models informing the participation strategies of renewable power producers. In our model, renewables optimize the forecast against the day-ahead and real-time prices, thereby maximizing the average profits across the day-ahead and real-time markets. By doing so, they also implicitly enhance the temporal cost coordination of day-ahead and real-time markets. We base the equilibrium analysis on the theory of variational inequalities, providing results on the existence and uniqueness of regression equilibrium in energy-only markets. We also devise two methods to compute regression equilibrium: centralized optimization and a decentralized ADMM-based algorithm.

3 - Enre Best Publication Award - Energy: Scaling up Electric-Vehicle Battery Swapping Services in Cities: a Joint Location and Repairable-Inventory Model

Wei Qi, Tsinghua University, Beijing, China, People's Republic of

Battery swapping for electric vehicles is reviving but faces a tension between maximizing battery proximity to users and grid constraints on decentralized charging. Cities are adopting infrastructure where decentralized swapping stations replenish charged batteries from centralized charging stations. This paper models this network (handling non-Poisson swaps and stochastic operations) and proposes a joint location-inventory model with a nonconvex objective, solved exactly via constraint generation and parameter search algorithms (speeding up by three orders of magnitude over Gurobi). The key insight is that it is the faster charging at centralized stations, not just pooled demand, that justifies the “swap-locally, charge-centrally” network by reducing system-wide battery stocks. In a broader sense, this work deepens our understanding of how mobility and energy are coupled toward enabling smart cities.

4 - Enre Best Publication Award - Natural Resources: from Curtailed Renewable Energy to Green Hydrogen: Infrastructure Planning for Hydrogen Fuel-Cell Vehicles

Long He, George Washington University, Washington, DC, United States

Hydrogen fuel-cell vehicles (HFVs) are a promising solution for sustainable transportation, particularly in regions with excess renewable energy. They can simultaneously reduce energy curtailment and promote green transport. This paper addresses key challenges to HFV adoption—uncertain demand, lack of infrastructure, and a spatial mismatch between hydrogen production and consumption. We develop a planning model to determine optimal locations and capacities for refueling stations and production plants, along with necessary grid upgrades. Applying our model to China's Sichuan province, we find that the optimal infrastructure layout depends on the adoption target. Higher targets shift deployment to major transportation hubs. We also show that promoting HFVs helps reduce hydropower curtailment and that hydrogen's transport flexibility allows for greater cost savings and curtailment reduction than other alternative fuel vehicles, e.g., electric vehicles.

5 - Enre Best Publication Award - Environment and Sustainability: Flattening Energy-Consumption Curves by Monthly Constrained Direct Load Control Contracts

Ali Fattahi, Johns Hopkins University, Baltimore, MD, United States

Balancing electricity demand and supply is one of the most critical tasks that utility firms perform to maintain grid stability and reduce system cost. Demand-response programs are among the strategies that utilities use to reduce electricity consumption during peak hours and flatten the energy-consumption curve. Direct load control contracts (DLCCs) are a class of incentive-based demand-response programs that allow utilities to assign “calls” to customer groups to reduce their energy usage by a

prespecified amount for a given length of time. Given the rapid expansion of such contracts, we develop an integer stochastic dynamic optimization problem for executing DLCCs that minimizes total system cost subject to monthly and annual constraints on the number of times and hours customers can be called. We develop a hierarchical approximation approach, which consists of an annual problem and monthly problems. Motivated by the practice in a large utility firm in California, we incorporate a reduce-to-threshold policy that attempts to flatten energy-consumption curves whenever demand exceeds a given threshold. We verified the quality of our proposed approach on real data from the California Independent System Operator, which is the umbrella organization of the utility firms in California, and measured the quality of our solution against a lower bound. A large utility firm in California implemented our model and informed us that the additional reduction in cost was approximately 4%. Our sensitivity analysis reports the impact of managerial concerns on some policies to enhance customer experience and provides insights for improving the features of DLCC contracts.

ME30

Building B Level 2 B204

Human-Centric Energy Markets

Invited Session

ENRE: Electricity

Chair: Bolun Xu, Columbia University, New York, NY, United States

Co-Chair: Liudong Chen, Columbia University, New York, United States

1 - Market Options for Improving Fairness and Reliability of Wind Farm Grid Integration

Robert Mieth, Rutgers University, Piscataway, NJ, United States

Off-shore wind power is projected to become a major supplier of electric energy in the near future due to the maturity of the technology, its decreasing cost, larger turbine sizes, and the advantage of accessing higher and more consistent wind speeds at sea compared to onshore wind. At the same time, off-shore wind turbines are harder to maintain because of uncertain metocean conditions and limited maintenance vessels. This makes ensuring the reliability of off-shore wind farms a risk both for the wind farm and the grid operator because derating a wind farm for maintenance under uncertain wind speeds and maintenance access with non-trivial planning constraints can significantly impact profits and energy supply. This risk has largely been overlooked by existing research which has mostly focused on risk from uncertain wind production. This talk discusses the modeling and analysis of maintenance planning for off-shore wind in a competitive electricity market. First, we discuss the impact of wind farm maintenance on electricity prices and farm profits and present both strategic and grid-serving maintenance planning tools. Building on these results we discuss fair and welfare-maximizing options for market-based and data-driven coordination between the grid and off-shore wind farms with maintenance constraints.

2 - Fair-over-Time Distributed Energy Resource Coordination

Johanna Mathieu, University of Michigan, Ann Arbor, MI, United States, Hannah Moring, Xavier Farrell

There are numerous strategies to coordinate distributed energy resources (DERs) to provide a variety of services to the power grid. DER coordination can affect resources and participants unequally, for example, by excessively degrading or curtailing particular DERs more than others. However, few DER coordination strategies explicitly take into account fairness. In this talk, we explore fairness metrics and their applicability to the DER coordination problem. In particular, we investigate metrics from machine learning to identify metrics that could be incorporated into DER coordination problems and we summarize fairness metrics that have been used in the power systems literature. A key challenge is that most existing fairness metrics are static – ensuring fairness at a point in time. DER coordination problems are inherently dynamic and we often care about fairness over time, not at each time. The machine learning literature offers some ways to think about fairness over time and, more generally, how to incorporate fairness into dynamic power systems problems. We use a specific DER coordination problem – the problem of computing dynamic operating envelopes – to demonstrate how incorporation of a fair-over-time metric changes DER coordination solutions, and highlight the trade-offs that arise.

3 - A Prudent Framework for Understanding Risk-Awareness in Demand Response

Liudong Chen, Columbia University, New York, NY, United States, Bolun Xu

We show that risk-aware behaviors in demand response originate from superquadratic state-dependent cost functions and price uncertainty with skewed distributions. We obtain such results through developing a novel theoretical demand response framework that combines non-anticipatory multi-stage decision-making with superquadratic cost functions. We introduce the concept of prudent demand, defined by a positive third-order derivative of the cost function, which is the first principle for risk-averse behavior despite a risk-neutral objective. Our analysis establishes that future price uncertainty affects immediate consumption decisions, and the extent of this response scales proportionally with the skewness of the price distribution. We visualize our theoretical findings through numerical simulations and illustrate their practical implications using a real-world case study.

4 - End-to-End Learning for Reliable Energy System Operation

Nicolas Christianson, California Institute of Technology, Pasadena, CA, United States

Machine learning (ML) can significantly improve efficiency in energy system operation. However, conventional ML approaches come with no a priori guarantees on reliability, and thus pose potential risks when deployed in the real world. In this talk, I will discuss recent work developing an end-to-end framework for training ML models for critical energy applications that are reliable by design. I will focus primarily on the application of training classification models for power grid contingency screening with provable guarantees on false negative rate - i.e., guarantees ensuring that no unsafe dispatch scenarios are misclassified as safe, which are critical for maintaining system reliability. I will also highlight how this framework can be extended to broader notions of reliability, such as enforcing market participants' varied risk preferences. In both cases, our end-to-end approach enables significant, data-driven performance improvements while providing provable guarantees on reliability.

5 - Optimal Pricing for Equitable Energy Access in Energy Communities

Siying Li, Cornell University, Ithaca, NY, United States, Timothy Mount, Lang Tong

We consider the design of an optimal pricing mechanism in an energy community consisting of members with diverse socioeconomic backgrounds, including varying income levels and differing capacities to access distributed energy resources such as solar power and storage systems. As part of the regulated retail market in a distribution system, the energy community is operated by a non-profit operator that sends pricing signals to its members, and each member maximizes individual consumption surplus. We formulate a bilevel optimization where the operator sets the pricing policy to maximize community welfare subject to constraints that (i) the community meets the desired equity level defined by the Rawlsian equity measure, and (ii) each community member has a higher consumption surplus within the community than being outside the community under the regulated net energy metering tariff. We show that an optimal pricing policy is the Dynamic Net Energy Metering with a discriminative connection charge, which achieves all points on the Pareto front of the equity-community welfare tradeoff. We show further that the optimal pricing with a higher Rawlsian equity must Lorenz-dominate one with a lower Rawlsian equity.

ME31

Building B Level 2 B205

Analytics for Cybersecurity Risk and Supply Chain Disruption

Invited Session

Decision Analysis Society

Chair: Youngsoo Kim, Loyola University Chicago, Chicago, United States

1 - Decoding Diversity: Whether, when, and Why Female Directors Enhance It Governance

Anqi Wu, Florida International University, Miami, FL, United States, Sebastian Schuetz, Xueping Liang

This study addresses the ongoing debate about the impact of female directors on corporate governance. While skepticism persists regarding the value women bring to corporate boards, and mixed evidence complicates the picture, this research examines whether, when, and why female directors make unique contributions in the realm of IT governance and data breach risks. By integrating arguments from prior literature, research on gender differences, and critical mass theory, we theorize the mechanisms explaining when and why female directors should make a difference. We test our hypotheses using a dataset from firms between 2011 and 2021. Our findings indicate that female directors reduce firms' data breach risks by 7.94%, with the effect being driven by their lower risk preferences and more participative leadership style. We also find that female directors need to reach a critical mass on corporate boards (i.e., $n > 2$) for these effects to materialize. These results are robust to alternative specifications and explanations. Collectively, our study challenges prevailing skepticism against female directors by highlighting the unique contributions—lower risk preferences and participative leadership style—that female directors bring to IT governance. Given the ongoing under-representation of women on corporate boards, our findings inform director recruitment practices by supporting the case for more equitable representation.

2 - Supply Chain Cyberattacks and Their Stock Market Impacts on Downstream Firms

Shreya Gokhale, University of Illinois Urbana-Champaign, Champaign, IL, United States, Gopesh Anand, Abhijeet Ghoshal, Yang Gao

Supply chain cyberattacks exploit vulnerabilities in upstream suppliers to compromise downstream customer firms. Unlike cyberattacks that directly target firms, supply chain cyberattacks infiltrate trusted supplier-customer relationships, causing widespread disruptions. While extensive research has examined the effects of cyberattacks on directly targeted firms, little is known about the impact of supply chain cyberattacks on downstream firms reliant on compromised suppliers. Given the increasing sophistication and the growing numbers of such attacks, understanding their financial consequences is crucial. This study examines whether supply chain cyberattacks affect the stock market performance of downstream firms. Using a novel dataset of supply chain cyberattacks between 2020 and 2023, we employ event study analyses to measure cumulative abnormal returns for downstream firms following cyberattacks on their suppliers. Our findings reveal a significant negative impact on the stock market performance of downstream firms, indicating that these cyberattacks extend beyond directly targeted suppliers to disrupt the supply chain. We further investigate the underlying mechanism and find that the supply chain contagion effect - which suggests that disruptions at compromised suppliers propagate downstream - plays a more dominant role than the blame attribution effect - which suggests that investors penalize downstream firms for failing to oversee their suppliers' cybersecurity practices. Additionally, we find that the negative stock market impact is more pronounced for downstream firms with weaker bargaining power, lower technical expertise, and explicit mentions in news reports. These findings underscore the importance of supply chain cyberattacks and provide actionable insights for firms to mitigate financial risks arising from these attacks.

3 - Innovation and Reputational Risk In Global Supply Chains

Kedong Chen, Rensselaer Polytechnic Institute, Troy, NY, United States

Despite growing regulatory pressure and corporate commitments to ethical sourcing, labor rights violations remain pervasive across global supply chains. Existing governance approaches, such as audits and certification schemes, have shown limited success in addressing these persistent challenges. Grounded in the relational capital perspective, this study intends to examine how buyers' process innovations, particularly those leveraging Artificial Intelligence (AI), can contribute to improving suppliers' labor practices and mitigating reputational risks. For instance, AI-driven demand forecasting in Unilever has stabilized supplier order volumes, enabling suppliers to reduce overtime and invest in workforce development, while in Walmart with its AI-powered inventory system reduced demand variability, allowing suppliers to better plan production and improve labor conditions. To account for heterogeneity in the relationships, the study further examines how the strength of buyer-supplier relationships and inter-firm power asymmetry moderate the relationship between process innovations of buyers and suppliers' labor compliance.

4 - Impact of Ransomware on Emergency Medical Services: Empirical Estimation and Risk Assessment Framework

Pouria Arsalani, THE UNIVERSITY OF ALABAMA, Tuscaloosa, AL, United States, Nickolas Freeman, Iman Dayarian, Youngsoo Kim, Xiaosong Peng, Ming Zhao, Jamie Gray, William Crawford

The healthcare industry has increasingly become a major victim of cyberattacks called ransomware, which poses nontrivial challenges to the sector because ransomware can disrupt their operations rather than merely breach data. A ransomware attack on a hospital often forces patients to be rerouted away from the closest facility, which can severely deteriorate the operations of emergency medical services (EMS). Utilizing a unique dataset that captures two ransomware breaches at large hospitals in Alabama, we empirically examine the impact of ransomware on ambulance patient transit, which is an integral part of EMS. We find that one ransomware breach significantly deteriorates the performances in three main segments of EMS operations, but the other ransomware breach does not. A further investigation suggests that the contrasting results between the two cases can be attributed to the network of available hospitals around the breached hospital. Motivated by our empirical findings, we develop a simulation model framework to assess the impact of ransomware attacks on EMS systems under various breach scenarios. We apply this framework to show that isolating hospital IT systems can substantially reduce operational disruptions during ransomware breaches and to explore a range of factors that shape EMS system susceptibility to ransomware disruptions. Overall, our findings can inform policymakers and healthcare administrators about the adverse effects of ransomware on EMS operations and highlight the importance of assessing the resilience of regional hospital networks, which can help develop more targeted and collaborative strategies to mitigate ransomware disruptions.

5 - Cybersecurity Starts at Home: Modeling Parental Behavior with Com-B

Ning Yang, Loyola University Chicago, Chicago, IL, United States, Inaiya Armeen

Despite growing awareness of online threats, children's digital activities have become an overlooked vector of cybersecurity risk. This study introduces a behaviorally informed framework to understand parental roles in managing and mitigating these risks, applying the COM-B (Capability, Opportunity, Motivation, Behavior) model to analyze protective behaviors within the household. We took a mixed-method approach, including 33 parent-child online interviews and surveying over one hundred U.S. parents, to assess psychological readiness, resource availability, and motivational drivers of digital safety practices regarding domestic devices and the environment. The results from the structural equation modeling show that motivation and opportunity significantly predict parental protective behavior, while capability exhibits an indirect impact on parental behaviors via motivation. Parents are often aware of cyber risks but lack the tools, time, or social support to sustain protection behaviors. Our studies indicate that parents should stop considering household cybersecurity as a technical problem requiring knowledge and skills. Instead, they need to approach this issue as a behavioral risk management challenge and actively seek interventions to strengthen their motivation and build resilience to shield their children from evolving digital threats. To support this, schools and communities must play a greater role in equipping parents with the resources, guidance, and social reinforcement needed to succeed.

ME33

Building B Level 2 B207

Seth Bonder Session

Award Session

Health Applications Society

Chair: Arielle Anderer, Cornell University, Johnson School, Ithaca, NY, United States

1 - Improving Access to Essential Medicines in the Sierra Leone via Machine Learning

Tsai-Hsuan (Angel) Chung, The Wharton School, UPenn, Philadelphia, PA, United States, Jatu Abdulai, Patrick Bayoh, Lawrence Sandi, Francis Smart, Hamsa Bastani, Osbert Bastani

A critical challenge in healthcare systems in Low- and Middle-Income Countries (LMICs) is the efficient and equitable allocation of scarce resources, particularly essential medicines. This problem is complicated by limited high-quality data, which restricts the applicability of traditional data-driven techniques. We propose a novel machine learning framework for essential medicines allocation, which leverages a combination of multi-task learning and decision-aware learning to improve sample efficiency and ensure equitable allocation. In collaboration with the Sierra Leone national government, our framework has been deployed nationwide as a decision support tool to help reduce waste and improve essential medicines allocation. Our evaluation using synthetic difference-in-differences analysis demonstrates a 19% increase in medicine consumption, with no changes to the supply, improving access for approximately 3.7 million women and children under five. Through experimental validation,

we demonstrate that our approach also significantly outperforms baseline approaches. Our work demonstrates the tangible impact of machine learning in optimizing high-stakes decisions in resource-constrained settings, improving efficiency while ensuring equity and cost-effectiveness.

2 - Optimization, Simulation and Predictive Models to Improve Response TIMES in Out-of-Hospital Cardiac Arrest

Marziehsadat Rezaei, University of Louisville, Louisville, KY, United States, Monica Gentili, Evan Kuhl

Over 350,000 out-of-hospital cardiac arrests (OHCA) occur annually in the U.S., with survival rates as low as 8–10%. Early defibrillation within 3–5 minutes can raise survival above 50%, highlighting the need for timely intervention. Delays often stem from Emergency Medical Service (EMS) operational limitations and challenging geographic access. Unmanned Aerial Vehicles (UAVs), or drones, present a promising solution to overcome these barriers by rapidly delivering Automated External Defibrillators (AEDs) to the scene. Prior studies suggest drones reduce response times and improve outcomes, yet many rely on i) narrow settings (e.g., regional focus), ii) simplified assumptions (e.g., fixed OHCA rates), and iii) lack of EMS coordination. My research aims at filling these critical gaps by designing and evaluating a coordinated drone-EMS AED delivery system using predictive, optimization and simulation models to improve emergency response time to OHCA. I have begun preliminary OHCA risk mapping and EMS data collection in Jefferson County as the initial testbed, but the approach will be scalable and generalizable to be applied to other regions. My research is structured around three objectives. Objective 1 develops optimization models for the optimal design of a coordinated drone-EMS AED delivery system. Objective 2 develops predictive and survival models to evaluate the impact of different delivery strategies. Objective 3 builds a simulation model to assess the performance of the coordinated system. The coordinated system will be compared with the current EMS system in terms of (1) survival outcomes, (2) response times, (3) resource utilization, and (4) scalability and cost-effectiveness.

3 - Context-Aware Optimization Models for Improving Population-Level Healthcare Delivery

Abel Sapirstein, Georgia Institute of Technology, Atlanta, GA, United States, Lauren Steimle, Mathieu Dahan, Meghan Meredith, Geir Godager

Outcomes for non-communicable diseases, such as cancers, heart disease, and diabetes, have significantly improved over the last half century due to medical advances, such as improved screening and biopsies. However, systemic barriers, including limited geographic access, low patient awareness, and poor cooperation between health systems planners, prevent some populations from receiving needed care. In this talk, I will discuss my proposed research agenda to address three practically motivated problems from different global contexts: access to cancer care in Rwanda, regionalization of maternity care in the US, and capacity building in long-term care facilities in Norway. These problems require the relaxation of common modeling assumptions surrounding provider and patient behaviors. As a result, we formulate novel models that require new solution approaches to solve. More specifically, in Rwanda, we model investments in the road network and cancer centers simultaneously under uncertainty in patient demand and the impact of seasonal weather events. We use stochastic optimization and develop a purpose-built branch-price-and-cut algorithm and approach optimality. In our ongoing work on maternal health in the US, we relax the assumption that patients always seek care at their closest facility to capture patients' bypassing behaviors. Finally, I'll share our planned work in Norway, in which we will model sustainable collaborative decision making, relaxing the assumption of a single, central planner. This work will consider uncertainty in disease costs and burdens when constructing long-term care facilities with large fixed costs. I will outline how this research agenda is aligned with my longer-term career goals.

ME34

Building B Level 2 B208

Statistical Learning and Data-driven Decision-making

Invited Session

OPT: Machine Learning

Chair: Kaizheng Wang, Columbia University, New York, NY, United States

1 - Surprises in Mean Estimation and Regression over Lp Balls

Reese Pathak, UC Berkeley, Berkeley, CA, United States

Empirical risk minimization and its regularized counterparts are the optimization workhorses of modern machine learning and statistical practice. On the other hand, despite years of study, the performance of these methods is poorly understood in even the simplest scenarios. We discuss new results on empirical risk minimization in mean estimation and linear regression setups where the underlying parameter is assumed to lie in an ℓ_p ball. Depending on the signal to noise ratio, we characterize the rate-suboptimality or optimality of some standard methods: empirical risk minimization (ERM), regularized ERM, and least-norm interpolation. These results point to the type of shrinkage that is required to attain rate optimality in these problems. We discuss the consequences of these results for modern learning setups which may entail distribution shift or learning in the presence of low noise.

Based on multiple papers, joint work with Liviu Aolaritei (UC Berkeley), Michael I. Jordan (UC Berkeley & INRIA), Gil Kur (ETH Zurich), Cong Ma (U. Chicago), and Annie Ulichney (UC Berkeley).

2 - Uncertainty Quantification for LLM-Based Survey Simulations

Chengpiao Huang, Columbia University, New York, NY, United States, Yuhang Wu, Kaizheng Wang

We investigate the use of large language models (LLMs) to simulate human responses to survey questions, and perform uncertainty quantification to gain reliable insights. Our approach converts imperfect LLM-simulated responses into confidence sets for population parameters of human responses, addressing the distribution shift between the simulated and real populations. A key innovation lies in determining the optimal number of simulated responses: too many produce overly narrow confidence sets with poor coverage, while too few yield excessively loose estimates. To resolve this, our method adaptively selects the simulation sample size, ensuring valid average-case coverage guarantees. It is broadly applicable to any LLM, irrespective of its fidelity, and any procedure for constructing confidence sets. Additionally, the selected sample size quantifies the degree of misalignment between the LLM and the target human population. We illustrate our method on real datasets and LLMs.

3 - A Statistical Theory of Overfitting for Imbalanced Classification

Kangjie Zhou, Columbia University, Philadelphia, PA, United States, Jingyang Lyu, Yiqiao Zhong

Classification with imbalanced data is a common challenge in data analysis, where certain classes (minority classes) account for a small fraction of the training data compared with other classes (majority classes). Classical statistical theory based on large-sample asymptotics and finite-sample corrections is often ineffective for high-dimensional data, leaving many overfitting phenomena in empirical machine learning unexplained.

In this paper, we develop a statistical theory for high-dimensional imbalanced classification by investigating support vector machines and logistic regression. We find that dimensionality induces truncation or skewing effects on the logit distribution, which we characterize via a variational problem under high-dimensional asymptotics. In particular, for linearly separable data generated from a two-component Gaussian mixture model, the logits from each class follow a normal distribution $\mathcal{N}(0,1)$ on the testing set, but asymptotically follow a rectified normal distribution $\max\{\kappa, \mathcal{N}(0,1)\}$ on the training set—which is a pervasive phenomenon we verified on tabular data, image data, and text data. This phenomenon explains why the minority class is more severely affected by overfitting. Further, we show that margin rebalancing, which incorporates class sizes into the loss function, is crucial for mitigating the accuracy drop for the minority class. Our theory also provides insights into the effects of overfitting on calibration and other uncertain quantification measures.

4 - Transfer Learning of Cate with Kernel Ridge Regression

Seok-Jin Kim, Columbia University, New York, NY, United States, Hongjie Liu, Molei Liu, Kaizheng Wang

The proliferation of data has sparked significant interest in leveraging findings from one study to estimate treatment effects in a different target population without direct outcome observations. However, the transfer learning process is frequently hindered by substantial covariate shift and limited overlap between (i) the source and target populations, as well as (ii) the treatment and control groups within the source. We propose a novel method for overlap-adaptive transfer learning of conditional average treatment effect (CATE) using kernel ridge regression (KRR). Our approach involves partitioning the labeled source data into two subsets. The first one is used to train

candidate CATE models based on regression adjustment and pseudo-outcomes. An optimal model is then selected using the second subset and unlabeled target data, employing another pseudo-outcome-based strategy. We provide a theoretical justification for our method through sharp non-asymptotic MSE bounds, highlighting its adaptivity to both weak overlaps and the complexity of CATE function. Extensive numerical studies confirm that our method achieves superior finite-sample efficiency and adaptability. We conclude by demonstrating the effectiveness of our approach using a 401(k) eligibility dataset.

5 - Causal Inference with High-Dimensional Discrete Covariates

Zhenghao Zeng, Carnegie Mellon University, Pittsburgh, PA, United States, Sivaraman Balakrishnan, Yanjun Han, Edward Kennedy

When estimating causal effects from observational studies, researchers often need to adjust for many covariates to deconfound the non-causal relationship between exposure and outcome, among which many covariates are discrete. The behavior of commonly used estimators in the presence of many discrete covariates is not well understood since their properties are often analyzed under structural assumptions including sparsity and smoothness, which do not apply in discrete settings. In this work, we study the estimation of causal effects in a model where the covariates required for confounding adjustment are discrete but high-dimensional, meaning the number of categories d is comparable with or even larger than sample size n . Specifically, we show the mean squared error of commonly used regression, weighting and doubly robust estimators is bounded by $d^2/n^2 + 1/n$. We then prove the minimax lower bound for the average treatment effect is of order $d^2/n^2 \log^2 n + 1/n$, which characterizes the fundamental difficulty of causal effect estimation in the high-dimensional discrete setting, and shows the estimators mentioned above are rate-optimal up to log-factors. We further consider additional structures that can be exploited, namely effect homogeneity and prior knowledge of the covariate distribution, and propose new estimators that enjoy faster convergence rates of order $d/n^2 + 1/n$, which achieve consistency in a broader regime. The results are illustrated empirically via simulation studies.

ME35

Building B Level 2 B209

Disrupting Human Trafficking

Invited Session

Public Sector OR

Chair: Daniel Lopes da Silva, Florida Polytechnic University, Lakeland, FL, United States

1 - Identifying the Digital Fingerprint of Illicit Massage Businesses

roya shomali, The University of Alabama, Tuscaloosa, AL, United States, Nickolas Freeman, Gregory Bott, Iman Dayarian

Illicit massage businesses (IMBs) masquerade as legitimate massage parlors while facilitating illicit activities, including sex trafficking. Traditional detection methods, which often rely on online reviews, fall short in revealing these covert operations. This research investigates how cell phone mobility data can be leveraged to uncover the digital fingerprint of IMBs by designing features that describe these businesses' operation and demand patterns. IMBs are key nodes in human trafficking networks, posing significant risks to exploited individuals and challenging law enforcement. A deeper understanding of mobility patterns and operational anomalies can enhance surveillance capabilities and inform targeted intervention strategies by governmental and non-profit organizations. Our study integrates large-scale mobility datasets with a well-delineated feature design process that spans temporal, spatial, and anomalous traffic dimensions. We employ a factorial experimental design to evaluate feature granularity and interaction effects and develop an XGBoost classifier to distinguish illicit from legitimate massage businesses. We apply our model to a nationwide holdout dataset to identify illicit businesses and delineate their spatial clusters, thereby pinpointing hot spots.

The proposed approach demonstrates robust performance across evaluation metrics, highlighting the framework's ability to flag IMBs across varied geographic contexts. Key features—assessed at specific granularity levels, including defined temporal visit patterns, catchment distance metrics, and elevated off-hours activity—emerge as strong indicators of illicit behavior. The digital fingerprints and predictive insights we derive provide law enforcement and social service agencies with actionable intelligence, enabling more effective prioritization of investigations and targeted interventions to disrupt illicit trafficking networks.

2 - Minimum Cost Flow Interdiction with Shifting Capacity for Human Trafficking Networks

Frederick Miller, Northeastern University, Somerville, MA, United States, Kayse Maass

Traditional network interdiction problems assume that the interdictor is the player that affects the network's performance. In reality, followers may adapt to interdictions by changing certain aspects of the network. This behavior is representative of applications in human trafficking (HT), where traffickers aim to respond to interdictions to recover lost supply or revenue. We present a multiperiod max - min cost flow network interdiction problem in which the graph capacities have the ability to shift over time. Specifically, we model the follower's ability to change the capacities of supply and demand nodes or arc capacities in response to the round's interdictions. This allows the follower's flow network to become more resilient to interdiction, making the interdictor's decisions less effective than it was in previous rounds. To model this, we developed a bilevel mixed integer nonlinear optimization problem. We study how supply and demand shift to nearby victims, showing how interdiction impacts those that remain in the network.

3 - Robust Link Prediction Models for Transnational Illicit Trafficking Networks

Hasini Balasuriya, University of Louisville, Louisville, KY, United States, Cagla Acun, Monica Gentili

Illicit operations, such as human and organ trafficking, pose a serious global threat. These operations are designed strategically to avoid detection by law enforcement, using decentralized structures and hidden communication channels, which makes them difficult to analyze and disrupt. Furthermore, the data available on these types of operations is often highly sparse and incomplete, adding another layer of complexity to their study.

To address these challenges, our research applies novel link prediction models to identify interactions or connections that may be missing in the available data. The analysis is based on datasets provided by external collaborators, including web-scraped news articles and online sources processed using large language models to extract relevant information. We introduce two separate, tailored, and robust link prediction models aimed at uncovering hidden links of both human and organ trafficking networks. To assess the strength of our models, we explore their robustness under different levels of missing data. By examining how the models perform under simulated scenarios of extreme conditions, we aim to ensure their reliability when applied to similar real-world trafficking network data.

Beyond predicting hidden connections, our study includes structural analysis to find key hubs and influential nodes within these networks. This integrated view helps policymakers design more targeted and effective strategies to disrupt illicit networks and mitigate their societal impact.

4 - Strategic Disruption of Human Trafficking Networks: a Multi-Layer Network Science and Interdiction Optimization Framework

Riad Al Hasan Abir, OHIO UNIVERSITY, Athens, OH, United States

Human-trafficking investigations rely heavily on surveillance and survivor testimony. This produces fragmented views of criminal enterprises and exposes victims to traumatization. Using network analytics and optimization, the project aims to deliver actionable intelligence to law enforcement. The main goal is to optimally allocate resources on the most vulnerable points in trafficking networks, accelerating case closure while minimizing harm. It also offers a blueprint for cross-agency data-sharing protocols. This presentation proposes an integrated network-science and interdiction-modeling framework that brings together heterogeneous digital traces, financial transactions, phone records, ads, geolocation activity, and investigative interviews, into a dynamic, multi-layer network. This method allows us to capture flows of money, communication, and control; key evidential features in an investigation. Temporal evaluation of the networks are analyzed with centrality, k-core, and role-equivalence metrics to reveal resilient organizing kernels and hidden brokers. These metrics feed a stochastic bilevel max-flow interdiction model that determines which nodes or edges to investigate, freeze, or disable so the maximum amount of illicit funds and victim movement is blocked under a fixed investigative budget. Column-and-constraint generation and decomposition heuristics enable the solution of realistically sized enterprise graphs drawn from real human trafficking investigations. We produce a set of experiments that include budget sweeps, layer-specific strikes, and simultaneous versus sequential actions quantify trade-offs among disruption effectiveness, resource

expenditure, and impact on survivors. The main goal of the project is to produce prioritization dashboards, visualizations, and reproducible notebooks that support law enforcement agencies producing “what-if” strategies before deploying agents.

5 - From Darknet to Daylight: Mapping Technology’s Role in Trafficking Control

Wojtek Wolfe, Rutgers University, Camden, NJ, United States, Srinivasa Puranam

Human trafficking exploits the same digital networks that power global commerce. Social platforms, encrypted messaging, and cryptocurrency let recruiters reach vulnerable targets, control victims, and wash profits. The same technologies also arm investigators through machine learning pattern matching, transparent blockchain ledgers, and mobile reporting tools. This study maps that contest.

A systematic scan of academic, industry, and open source evidence tracks how online recruitment, surveillance, and illicit finance evolve. Case studies then test countermeasures, from artificial intelligence triage of classified ads to tamper resistant logistics monitoring. Data sources include web scraped social content, synthetic vulnerability models, and darknet market tracking. Privacy and victim safety guide every step.

Findings show impact depends on policy alignment. Blockchain pilots cut forced labor risk when buyers share data in real time, yet predictive policing stalls where privacy law blocks intelligence exchange. As firms reshore to stricter jurisdictions, trafficking pressure shifts toward regions with weaker oversight, deeper capital stress, and rising displacement.

The project delivers two outcomes. A diagnostic matrix ranks technical tools by readiness, scalability, and potential harm. A policy playbook links those ratings to procurement rules, transparency statutes, and multilateral data sharing accords. Aligning technology with legal and economic incentives can shrink the space in which traffickers operate even in increasingly fragmented markets.

ME36

Building B Level 2 B210

SpORts IV

Invited Session

spORts

Chair: Scott Nestler, University of Florida, Gainesville, FL, United States

1 - The Economics of Income Sharing for Athletes: An Insurance-Based Approach

Christopher Gaffney, Drexel University, Philadelphia, PA, United States

Income Sharing Agreements (ISAs) have emerged as a popular financial tool for early career athletes, allowing them to trade a percentage of future earnings for immediate capital. This creates a dual benefit: athletes gain near-term financial stability while investment firms capture value through calculated risk exposure. In this study, we develop a mathematical framework to analyze and optimize the structure of these agreements. We draw parallels between ISA variables such as advance size, repayment rates, and repayment maximums, and traditional insurance concepts such as coverage, deductible, and premium loading. Using these features, we derive Pareto-optimal agreement structures. This research contributes to the growing literature on innovative sports financing mechanisms and provides actionable insights for athletes, agents, and financial institutions in the sports industry.

2 - Comparing Nba and Wnba Fan Reactions: a Topic Modeling Analysis of X (Twitter) Discussions

Scott Nestler, University of Florida, Gainesville, FL, United States

This study compares how NBA and WNBA fans discuss games on Twitter, focusing on differences in player versus team emphasis and emotional tone. Using Latent Dirichlet Allocation (LDA) topic modeling, we analyzed 50,000 tweets collected during the 2023 NBA and WNBA Finals. The analysis is grounded in social identity, arousal, and expectancy violation theories to interpret fan loyalty, emotional engagement, and reactions to unexpected outcomes. Data were preprocessed with tokenization, lemmatization, and emoji conversion. Five main topics were identified for each league. NBA discussions were predominantly player-focused, with frequent mentions of stars like Jamal Murray, Jimmy Butler, and Nikola Jokic. In contrast, WNBA conversations centered more on teams, especially the Las Vegas Aces and New York Liberty, with terms like “A’ja Wilson,” “team,” and “win” appearing often. Player mentions in NBA tweets were 40% higher than in WNBA tweets, while team dynamics comprised 35% of WNBA discussions but only 15% in NBA tweets. Sentiment analysis revealed that 65% of WNBA-related tweets were positive, compared to 38% for NBA tweets, which also contained more negative terms. NBA fan conversations more frequently referenced specific game moments, such as scoring and rebounds. These findings highlight distinct cultural and structural patterns in fan engagement across the two leagues. Implications for social media managers and content creators aiming to engage NBA and WNBA audiences will be discussed.

3 - Creating a Crystal Ball for English Premier League Clinch and Relegation

Alexandra Newman, Colorado School of Mines, Golden, CO, United States

The English Premier League is one of soccer's most competitive professional sports leagues. We discuss magic numbers of interest, e.g., avoiding relegation, and how we construct mixed-integer programming models to produce these numbers. We compare our results against those available in the mainstream media, and showcase a website with said numbers.

4 - Variations in Player Valuations in Expanded and Contracted Labor Markets: Evidence from Indian Premier League Auctions

Sachin Narayanan, Florida State University, Tallahassee, FL, United States

Entering its 18th season, the financially lucrative Indian Premier League (IPL) ranks only behind the NFL in per-game revenue (Jamkhandikar & Ganguly, 2022). IPL teams build their roster annually through single-item English auctions, bidding on players based on talent and team needs. Auctions occur in two forms: a mega auction every 3-4 years, with more roster spots, and an annual mini auction to fill unexpected vacancies. This introduces a dynamic labor market for cricketers that expands (mega) and contracts (mini), potentially altering player valuations. Neoclassical economics propose that wages reflect supply and demand (Hicks, 1963). Expanding demand with unmatched supply tends to raise wages; contracting demand may decrease them (Rebitzer, 1995). The dynamic nature of the IPL auction, coupled with the criticism surrounding disproportionate wage distributions, raises the following questions about how firms value labor under varying market conditions: RQ1: How does a player’s Marginal Revenue Product (MRP) compare to their compensation under expanded and contracted labor markets? RQ2: Do factors affecting player wages (valuation) vary between auction types? Methodology Using advanced player/team performance (espnricinfo.com), auction (IPLT20.com) data, and team valuation reports (Forbes) we estimate MRP models (Humphreys & Pyun, 2017) for auctioned players, compare MRP with wages, and regress wages (MRP) on performance factors. Findings and Implications Findings indicate higher wages in mega auctions, positional preference for offense in mega, and overcompensation for veterans in mini auctions. These findings can guide fairer player pricing, identify undervalued talent, and inform efficient bidding strategies, offering insights applicable to other dynamic labor markets.

ME37

Building B Level 2 B211

Learned Optimizers and Decision-Focused Learning

Invited Session

OPT: Optimization Under Uncertainty

Chair: Paul Grigas, University of California, Berkeley, Berkeley, CA, United States

Co-Chair: Wyame Benslimane, University of California, Berkeley, Berkeley, 94705, United States

1 - Self-Supervised Penalty-Based Learning for Robust Constrained Optimization

Wyame Benslimane, University of California, Berkeley, Berkeley, CA, United States, Paul Grigas

We introduce a novel self-supervised learning framework for parameterized constrained robust optimization, an important class of optimization problems. Inspired by the exact penalty method used to solve constrained optimization problems, we propose a self-supervised penalty-based loss function to learn an approximate optimal solution mapping for parametric robust optimization problems. Our approach offers significant advantages as it leverages the problem's inherent structure to train our approximation without requiring prior knowledge of optimal solutions, thus bypassing the need for data labeling. This is particularly beneficial as data labeling often relies on traditional solvers that tend to scale poorly for large scale optimization problems. Additionally, we adapt our framework for robust constrained combinatorial optimization problems by incorporating a surrogate linear cost over mixed-integer domains, and a smooth approximation thereof, into the final layer of the network architecture.

We evaluate our method on three applications: a continuous multidimensional knapsack problem, its combinatorial (discrete) variant, and an inventory management problem. Experimental results show that our approach achieves high-quality approximations with significantly faster inference times compared to traditional solvers. Moreover, we demonstrate that adjusting the penalty parameter allows for effective control of the trade-off between solution optimality and robust feasibility.

2 - Overworking or Outsourcing: Integrated Prediction-Optimization for Human-Centric Last-Mile Delivery Zone Assignment

Meng Qi, Cornell University, Ithaca, NY, United States, Haoran Guo, Wei Qi

As courier overtime and burnout are increasingly exacerbated by significant fluctuations in e-commerce demand, emerging crowd-sourced delivery provides a flexible solution to manage service peaks and protect the full-time workforce. When determining delivery zone assignments between couriers and gig workers, the process often involves an objective function that aims to balance operational costs against courier overtime. However, in two-step predict-then-optimize approaches, simply introducing a human-centric objective in the optimization stage may not be effective due to its misalignment with the prediction goal. To address this critical challenge and ensure courier welfare, we propose an integrated prediction-optimization (IPO) approach that leverages the problem structure to guide the construction of a decision-focused feature-based service time estimator. Theoretically, we demonstrate that a point estimator suffices, which is a property that typically does not hold when the objective function is nonlinear. It avoids learning a distributional prediction, which is more complicated but often required with nonlinear objective functions. We also utilize geometric structures to identify tractable cases, the resulting estimators achieve a bounded excess decision risk. More broadly, this paper demonstrates how IPO approaches can be leveraged for employee well-being alongside operational efficiency.

3 - Understanding Fixed Predictions via Confined Regions

Connor Lawless, Stanford University, Stanford, CA, United States, Tsui-Wei Weng, Berk Ustun, Madeleine Udell

Machine learning models are designed to predict outcomes using features about an individual, but fail to take into account how individuals can change them. Consequently, models can assign fixed predictions that deny individuals recourse to change their outcome. This work develops a new paradigm to identify fixed predictions by finding confined regions in which all individuals receive fixed predictions. We introduce the first method, ReVer, for this task, using tools from mixed-integer quadratically constrained programming. Our approach certifies recourse for out-of-sample data, provides interpretable descriptions of confined regions, and runs in seconds on real world datasets. We conduct a comprehensive empirical study of confined regions across diverse applications. Our results highlight that existing point-wise verification methods fail to discover confined regions, while ReVer provably succeeds.

4 - Diffusion Models for Uncertainty-Aware Decision-Focused Learning

Zihao Zhao, Georgia Institute of Technology, Atlanta, GA, United States, Linghai Kong, Kai Wang

Decision-Focused Learning (DFL) integrates predictive modeling and optimization by training predictors to optimize the downstream decision target rather than merely minimizing the prediction error. While existing DFL frameworks have achieved significant success using deterministic neural networks to characterize uncertain parameters, we show that point predictions are often insufficient to capture the intrinsic stochasticity of real-world phenomena. To address this challenge, we propose the first diffusion-based DFL approach, which trains a diffusion model to represent the distribution of the uncertain parameters, and optimizes the decision by solving a stochastic optimization with samples drawn from the diffusion model. We show that backpropagating through stochastic optimization is equivalent to a weighted score function formulation that can be used to train the diffusion model, which provides a simple yet effective algorithm for training diffusion models end-to-end. This new integration of diffusion models and stochastic optimization leads to a new DFL approach for training diffusion models. Lastly, we empirically show that our diffusion-based DFL approach consistently outperforms state-of-the-art baselines in decision quality.

ME38

Building B Level 2 B212

Optimization Under Uncertainty

Contributed Session

Contributed

Chair: Rui Xi, Industrial Engineering Department, University of Tennessee, Knoxville, 1203 Sutters Mill, Knoxville, TN, 37909, United States

1 - Multi-Region Competitive Pricing and Market Dynamics of Hurricane Insurance Under Climate Change Risks

Dahui Liu, Donghua University, Shanghai, China, People's Republic of, Linda Nozick, Meghan Millea, Jamie Kruse, Rachel Davidson, Junkan Li

Homeowner insurance plays a crucial role in mitigating the impact of hurricanes on households and regional economies. Pricing the insurance is challenging, as premiums must balance between covering large, localized payouts and remaining affordable. Excessive premiums may result many homeowners uninsured, while inadequate pricing threatens insurer solvency. Climate change intensifies hurricane severity, leading to greater losses, yet homeowners and insurers perceive these risks differently—homeowners often underestimate the impact, whereas insurers anticipate it to safeguard against insolvency.

This study develops a Cournot-Nash oligopoly model to explore multi-region pricing strategies in the hurricane insurance market. A heuristic-based search method is employed to find equilibrium prices under complex demand and cost structures. We further explore scenarios with varying homeowner beliefs in climate change, assessing impacts on insurance prices, reinsurance costs, market penetration, and insurer profitability.

Applying the model to coastal North Carolina, results show that insurers offer lower rates per expected loss in high-risk coastal areas. Increased competition from new entrants drives prices down and boosts insurance uptake, benefiting homeowners but pressuring insurer finances. Climate change escalates premiums and reinsurance costs, placing growing financial burdens on both insurers and homeowners. However, as more homeowners acknowledge climate risks, insurers experience reduced financial stress and can achieve higher profitability through adjusted pricing strategies. These findings underscore the critical interplay between market dynamics, climate perception, and insurance system resilience in hurricane-prone regions.

2 - Bayesian Optimization Under Functional Uncertainty: a Robust Approach in Function Space

Pouya Ahadi, Georgia Institute of Technology, ATLANTA, GA, United States, Kamran Paynabar

We propose Functional Robust Bayesian Optimization (FRBO), a new framework for optimizing under functional uncertainty, where the entire mapping from design inputs to responses may vary within a structured function space. Unlike prior methods that model scalar input or output uncertainty, FRBO assumes the true function lies within a smooth, norm-bounded region around an unknown reference. We model this ambiguity using an RKHS ball and derive a robust surrogate objective that accounts for both interpolation error and epistemic variability. This enables principled acquisition without reliance on posterior sampling, allowing for nonparametric robustness in

black-box settings. FRBO supports both scalar and functional-valued responses and applies to diverse domains such as optics and materials design. We provide theoretical guarantees on worst-case regret and demonstrate that FRBO consistently outperforms baseline methods by reducing functional regret across a range of uncertainty levels.

3 - Applying Uncertainty Quantification to Improve Plans for Place-Based Projects

Brent Houchens, Sandia National Laboratories, Albuquerque, NM, United States, Mika Armenta, Amanda Wachtel, Sandra Begay

A framework will be presented that facilitates efficient, objective assessments of project plans for place-based (PB) work, offering guidance on attributes that are particularly impactful or critical. This framework is designed to improve the outcomes of PB projects that receive technical assistance (TA), which is often provided by the national labs on projects such as improving energy infrastructure and resilience. Currently there isn't a well-established structure for guiding communities or executing such TA work. While every project has unique characteristics and challenges, there is a recognition that, at a high level, there is a common workflow to projects, and critical attributes throughout that can determine the success.

Unique to this framework is the use of uncertainty quantification (UQ) to predict sensitivities of project outcomes to input attributes across the project space – spanning community buy-in, resource evaluation, permitting, integration, funding mechanisms, operations and maintenance, workforce development, decommissioning, etc. This identifies the most sensitive input attributes and propagates them through the project model, determining when these need to be strengthened, and preventing future bottlenecks and single-point failures.

A case study will be presented demonstrating the capabilities of the framework. Potentially conflicting attribute assessments will be considered from both subject matter experts and community members. This will help identify low, high and conflicting value attributes and potential model-form error or input bias.

Sandia National Laboratories is managed and operated by NTESS under DOE NNSA contract DE-NA0003525.

4 - Data-Driven Planning of Renewable Distributed Generation Under Uncertainty via Distributionally Robust Optimization

Azadeh Farsi, University of Arizona, Tucson, AZ, United States, Canqi Yao, Jianqiang Cheng, Kai Pan

As the integration of renewable distributed generation (RDG) into modern power grids accelerates, it becomes increasingly critical to develop effective planning methodologies that enhance renewable energy utilization while mitigating adverse effects such as phase unbalance. In this talk, we propose a novel two-stage distributionally robust optimization (DRO) model for the optimal siting and sizing of RDG units. The model captures uncertainty in carbon emission pricing and the active power output efficiency of renewable generators through a data-driven ambiguity set. The objective is to minimize the total cost, including both the capital investment in RDG installations and the operational costs incurred over the planning horizon. To promote social equity, the model incorporates energy equity constraints to ensure fair energy access across all buses in the distribution network. Moreover, to address the computational challenge caused by the uncertainties in both the objective function and the constraints, we develop an innovative approximation framework based on the Partial Sample Average Approximation method. The effectiveness of the proposed approach is demonstrated through numerical experiments conducted on the IEEE 33-bus radial distribution system.

5 - Solving the Pollution Routing Problem with Hilly Topography

Yasel Costa, Zaragoza Logistics Center (ESG50985993), Zaragoza, Spain, David Lai, Tom van Woensel

This study addresses the pollution routing problem under a particular setting that includes irregular topography and speed optimization. We propose two solution approaches (approximate and exact) for solving modified Solomon benchmarks and newly constructed real-life problem instances. Results show the good performance of proposed methods and suggest several managerial insights.

6 - Cloud Computing Resource Provisioning Under Uncertainty

Melis Boran, Middle East Technical University, Ankara, Turkey, Secil Savasaneril, Demet Batur

Cloud computing allows users to flexibly control, customize, and scale computing resources without infrastructure investment. Although its adoption has grown since the 2000s, effective cloud cost management remains a challenge. Infrastructure-as-a-Service (IaaS) providers offer computing resources through various pricing plans, and users must decide how much resource to provision with the best pricing plan in advance to meet their workload requirements over a planning horizon. This decision is challenging due to the inherent uncertainty in workloads; the demand for computing resources may vary significantly in both scale and timing.

We address this resource provisioning problem by modeling it as a Markov Decision Process (MDP) aiming to derive a cost-minimizing policy. The solution of the model yields a dynamic optimal provisioning policy. For small-scale problems, we prove that the optimal policy is a state-dependent provisioning policy, which is monotone and non-increasing in the number of available resources. Large-scale problems are intractable due to the curse of dimensionality. Thus, we propose a simple yet effective heuristic policy. We benchmark its performance against the optimal policy in small-scale instances and against commonly used benchmark policies in larger-scale settings. The results show that the heuristic closely approximates the optimal policy in small instances and outperforms benchmark policies in more complex settings.

7 - Sequential Resource Allocation for Humanitarian Operations Using Approximate Dynamic Programming

Sina Ansari, Driehaus College of Business, DePaul University, Chicago, IL, United States, Masih Fadaki, Ahmad Abareshi, Paul Tae-woo Lee

In humanitarian and nonprofit operations, distributing aid such as food, shelter, and medical supplies becomes challenging in an online setting, where the future demand is unknown, since allocation decisions are made in real time as uncertainties unfold. Being overly conservative in allocating items at the beginning of the supply chain to save stock for fulfilling demand further down the supply chain increases the likelihood of unallocated items (waste). On the other hand, fully addressing the demand of nodes in the earlier stages of the supply chain may negatively impact the equity of the allocation policy, as downstream nodes may receive significantly fewer items in proportion to their demand. This study proposes a framework for modeling the sequential decisions involved in this online resource allocation problem as a Markov Decision Process (MDP). Given that the size of the state-action space can become very large for this problem, standard dynamic programming methods in the reinforcement learning domain reach their limits, so using Approximate Dynamic Programming (ADP) is a practical solution. In this study, two methods of measuring downstream uncertainty are proposed, and Policy Function Approximation (PFA) is used to develop an optimal allocation policy. Numerical results and the application of the proposed model to the Food Bank of Southern Tier in New York suggest a reasonable balance between maximizing efficiency (minimizing the waste of unallocated items) and ensuring an equitable allocation.

8 - A Decision-Support Model for Irrigation Planning Under Flash Drought and Flood Risks Using Remote Sensing and Stochastic Optimization

Rui Xi, University of Tennessee, Knoxville, TN, United States, Yuefeng Hao, Mingzhou Jin

Climate change has increased the frequency and intensity of flash floods and droughts, posing serious challenges for agricultural systems. This project develops a decision-making model to help manage these risks. It uses GRACE and MODIS satellite data to track large-scale water storage and vegetation activity. It also includes SWAT hydrologic model outputs to simulate local water processes. These inputs capture daily water availability and crop stress. A stochastic programming framework is used to represent uncertainty in the timing and intensity of extreme weather events. The model supports both long-term investment planning and short-term operational decisions, such as irrigation scheduling and water removal. The expected outcome is a flexible and practical tool that provides farmers and agricultural decision-makers with more accurate and timely information. We also expect the framework to be adaptable to other regions and climate-related challenges.

ME39

Building B Level 2 B213

Submodular Function Minimization: An Interplay between Discrete and Continuous Optimization

Invited Session

OPT: Integer and Discrete Optimization

Chair: Andrei Graur, Stanford University, 729 Escondido Road, Apt 110, Stanford, 94305-7573, United States

1 - Sparse Submodular Function Minimization

Andrei Graur, Stanford University, Stanford, CA, United States

The submodular function minimization (SFM) problem is a foundational combinatorial optimization problem that encompasses directed hypergraph minimum cut problems and computing the size of the largest set in the intersection of two matroids. Given access to an evaluation oracle for a submodular function defined on subsets of a ground set of n elements, the goal of SFM is to compute a minimizer (possibly approximate) with as few queries and additional computation as possible. While SFM is a well-studied problem, its complexity has not been settled. In particular, the state-of-the-art upper bounds on queries for algorithms with $\text{poly}(n)$ additional computation are $O(n^3)$ queries, while the lower bound is $\Omega(n \log n)$. Consequently, we ask the question of whether better runtimes are possible when the minimizer is known to be k -sparse. We title the problem as k -sparse SFM and answer the question in the affirmative, providing an algorithm which makes $O(n \text{poly}(k))$ queries. In this talk, I will introduce the k -sparse SFM problem and some of the tools that enable us to obtain improved runtimes. I will start by presenting some structural properties that arise due to the sparsity assumption in the context of SFM. I will then present how we leverage these properties to obtain an algorithm that finds a minimizer in $\text{poly}(k)$ rounds (each round taking n^2 queries). Finally, I will briefly sketch the technical tools that enable us to obtain our $O(n \text{poly}(k))$ algorithm.

2 - Quadratic Decomposable Submodular Function Minimization with Application in Pagerank Computation for Hypergraphs

Pan Li, Georgia Institute of Technology, Atlanta, GA, United States

Hypergraphs are powerful tools for modeling complex, higher-order interactions in real-world data. When each hyperedge is assigned a submodular weight function, we obtain a submodular hypergraph, capturing intricate dependencies beyond pairwise relationships. Solving min-cut problems on such structures leads to Decomposable Submodular Function Minimization (DSFM), a convex optimization framework that effectively handles these challenges. However, DSFM primarily addresses cut-based objectives. To model diffusion processes—like those captured by PageRank in standard graphs—we extend this framework to incorporate quadratic regularization, resulting in Quadratic Decomposable Submodular Function Minimization (QDSFM). In this talk, we briefly introduce QDSFM, highlighting its formulation that combines quadratic terms with squared Lovász extensions of submodular functions. We present efficient optimization strategies, including a novel dual formulation and double-loop algorithms that integrate Random Coordinate Descent or Alternating Projections in the outer loop with modified Min-Norm-Point or Frank-Wolfe methods in the inner loop. These algorithms come with provable linear convergence rates, ensuring both theoretical robustness and practical efficiency. Applications of QDSFM demonstrate its versatility and effectiveness. Notably, it enables a hypergraph-adapted PageRank algorithm for local partitioning, offering provable performance guarantees. Additionally, in semi-supervised learning scenarios, QDSFM-based methods have shown significant improvements in prediction accuracy compared to existing approaches.

3 - Minimum Cost Adaptive Submodular Cover

Blake Harris, University of Michigan, Ann Arbor, MI, United States, Hessa Al-Thani, Yubing Cui, Viswanath Nagarajan

Adaptive submodularity is a fundamental concept in stochastic optimization, with numerous applications such as sensor placement, hypothesis identification and viral marketing. We consider the problem of covering an adaptive-submodular function at minimum expected cost, which generalizes the classic set cover and submodular cover problems to the stochastic setting. We show that the natural greedy policy has an approximation ratio of $4 \cdot (1 + \ln Q)$, where Q is the goal value. In fact, we consider a significantly more general objective of minimizing the p^{th} moment of the coverage cost, and show that the greedy policy *simultaneously* achieves a $(p + 1)^{p+1} \cdot (\ln Q + 1)^p$ approximation guarantee for all $p \geq 1$. All our approximation ratios are best possible up to constant factors (assuming $P \neq NP$). Moreover, our results extend to the setting where one wants to cover *multiple* adaptive-submodular functions, for which we obtain the same approximation guarantees. Finally, we evaluate our algorithm empirically on instances of hypothesis identification and observe that it performs very well in comparison to a lower bound that we compute.

4 - Discrete and Continuous Difference of Submodular Minimization

Marwa El Halabi, Samsung AI Lab, Montreal, QC, Canada, George Orfanides, Tim Hoheisel

Submodular functions, defined on continuous or discrete domains, arise in numerous applications. We study the minimization of the difference of submodular (DS) functions, over both domains, extending prior work restricted to set functions. We show that all functions on discrete domains and all smooth functions on continuous domains are DS. For discrete domains, we observe that DS minimization is equivalent to minimizing the difference of two convex (DC) functions, as in the set function case. We propose a novel variant of the DC Algorithm (DCA) and apply it to the resulting DC Program, obtaining comparable theoretical guarantees as in the set function case. The algorithm can be applied to continuous domains via discretization. Experiments demonstrate that our method outperforms baselines in integer compressive sensing and integer least squares.

ME40

Building B Level 2 B214

Frontiers in Causality, Experimentation, and Machine Learning

Invited Session

Data Mining

Chair: Ruoxuan Xiong, Emory University, Atlanta, GA, United States

1 - Estimating Bounds on Selection Bias with Outcomes Measured on the Selected

Adam Glynn, Emory University, Atlanta, GA, United States, Elisha Cohen

Outcome tests are a method for detecting bias in selection procedures using data on selected units. We use a principal stratification approach to establish lower bounds on this bias for general outcomes. We show that the Knox et al. (2020) bound is sharp for a binary outcome, and weaken the assumptions required for that bound. We also show that the analogous bound for non-binary outcomes is not sharp and provide sharper lower bounds. We illustrate these methods with a re-analysis of the data in Anzia and Berry (2011) on the delivery of federal spending by male and female members of Congress. Using their data, and the covariates from their analysis, we find that the unsharpened lower bound estimates that roughly 4% of men elected, would not have been elected, had they been women. That estimate increases to roughly 20% when sharpening the lower bound using our method.

2 - Strategic Experimentation: A Data-Driven Approach to Theory and Practice

Molly Offer-Westort, University of Chicago, Chicago, IL, United States

On digital platforms, experiments—especially A/B tests—are often used to evaluate whether a pre-specified intervention improves outcomes relative to a pre-specified control condition. The interventions may represent a test of a theory or evaluation of a product or design feature. However, there may be multiple interventions that could be consistent with this test and a single intervention may only imperfectly reflect the underlying construct of interest. When experiments yield null results, how should we interpret them: as evidence against the theory or feature, or only against that specific realization of the intervention?

This question becomes especially salient in large-scale digital environments, where we can leverage rich data and adaptive experimentation to explore the space of possible interventions. In this paper, I propose a framework for theory-informed experimental design that integrates both theory testing and theory building. First, I illustrate how strategically designed experiments allow us to explore the "theory space" for more informative data driven hypothesis tests. Second, I extend this framework to theory building, addressing cases where minimal theoretical grounding exists. This approach is especially valuable in digital contexts where we have scale but possibly limited prior theory, and where there may be a large range of candidate interventions. Through designs with strategic assignment algorithms, we can improve statistical power, manage multiple testing concerns, and guide experiments toward effective interventions. I discuss relevant applications to social media experiments conducted over Facebook Messenger.

3 - Designing Staggered Rollout Experiments with Indefinite Effect Lengths

Weibin Mo, Purdue University, West Lafayette, IN, United States

We consider the experimental design of staggered rollouts, where the treatment effects since intervention can last for a long and even infinite number of periods. In this case, the two-way-fixed-effect regression with a fixed length of effect window is generally misspecified, but the treatment effect coefficients can still be consistent estimates of the cumulative treatment effect in long run. We quantify the bias-variance trade-off under a large-panel asymptotic regime, that is, the numbers of treatment units and periods tend to infinity proportionally. Under such an asymptotic regime, we prove the consistency of the misspecified cumulative effect estimate, and the optimality of linear symmetric staggered rollouts. We also obtain an optimal design of always-treated, never-treated, and linear-rollout units that minimizes the asymptotic mean-squared error of the cumulative effect estimate. As two extreme special cases, the “flat” design with all units either always-treated or never-treated over time is bias-optimal, while the “diagonal” design with all linear-rollout units is variance-optimal. This complements Xiong et al. (2023)’s design with the optimality of linear rollout and the trade-off under lag misspecification. The managerial implication of such a trade-off is that, the linear staggered rollout can incur bias due to long-lasting treatment effects, and such a bias can be offset in a long-run design (large panel) with additional always-treated and never-treated units.

4 - Model-Assisted Inference for Dynamic Causal Effects in Staggered Rollout Cluster Randomized Experiments

Xinyuan Chen, Mississippi State University, Mississippi State, MS, United States

Staggered rollout cluster randomized experiments (SR-CREs) are increasingly used for their practical feasibility and logistical convenience. These designs involve staggered treatment adoption across clusters, requiring analysis methods that account for an exhaustive class of dynamic causal effects, anticipation, and non-ignorable cluster-period sizes. Without imposing outcome modeling assumptions, we study regression estimators using individual data, cluster-period averages, and scaled cluster-period totals, with and without covariate adjustment from a design-based perspective, where only the treatment adoption time is random. We establish consistency and asymptotic normality of each regression estimator under a finite-population framework and formally prove that the associated variance estimators are asymptotically conservative in the Lowner ordering. Furthermore, we conduct a unified efficiency comparison of the estimators and provide practical recommendations. We highlight the efficiency advantage of using estimators based on scaled cluster-period totals with covariate adjustment over their counterparts using individual-level data and cluster-period averages. Our results rigorously justify linear regression estimators as model-assisted methods to address an entire class of dynamic causal effects in SR-CREs and significantly expand those developed for parallel-arm CREs by Su and Ding (2021, JRSSB) to accommodate a wider class of complex experimental settings with staggered randomization.

5 - Asymptotic Theory of the Best-Choice Rerandomization Using the Mahalanobis Distance

Yuhao Wang, Tsinghua University and SQZ, Haidian District, China, People's Republic of, Xinran Li

Rerandomization, a design that leverages pretreatment covariates to improve balance between treatment groups, has gained increasing attention in both theory and practice. Bruhn and McKenzie (2009) identify two common approaches: (i) rerandomizing until covariate imbalance is below a prespecified threshold, and (ii) generating multiple randomizations and selecting the one with the best covariate balance. This paper studies the second approach, best-choice rerandomization, for which theoretical foundations remain underdeveloped.

We focus on best-choice rerandomization using the Mahalanobis distance to measure covariate imbalance, a widely used criterion due to its invariance to affine transformations of covariates. Our analysis examines the large-sample repeated sampling properties of this design, allowing both the number of covariates and the number of tried randomizations to increase with sample size.

We show that under best-choice rerandomization, the asymptotic distribution of the difference-in-means estimator is more concentrated around the true average treatment effect than under complete randomization. Based on this, we propose large-sample confidence intervals that are both valid and shorter than those from standard randomized experiments.

Furthermore, we demonstrate that with a moderate number of covariates and a polynomially increasing number of randomizations, best-choice rerandomization can attain nearly optimal precision, approaching that achievable with perfectly balanced covariates. The proposed theory and methods are also illustrated with real field experiments, highlighting their practical relevance.

ME41

Building B Level 3 B301

Senior PhD Student Advice – Career Options in Healthcare

Panel Session

Health Applications Society

Co-Chair: Melike Yildirim, Wayne State University, Detroit, MI, United States

Co-Chair: Huaiyang Zhong, Virginia Tech, Blacksburg, VA, United States

1 - Moderator Panelist

Huaiyang Zhong, Virginia Tech, Blacksburg, VA, United States

This panel discussion will focus on careers outside of engineering and business schools at the intersection of IEOR and Healthcare. The panel will feature both academic faculty, from schools of public health and medical schools, and industry professionals who have obtained their terminal degree in IE/OR/MS, and will showcase their career experience. It will also feature advice for final-year PhDs students who are looking to enter the job market. Each panelist will introduce themselves and talk briefly about their career trajectory before opening the panel to questions from the audience.

2 - Panelist

Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States

3 - Panelist

Shalome Hanisha Tatapudi, University of South Florida, Tampa, FL, United States

4 - Moderator Panelist

Melike Yildirim, Wayne State University, Detroit, MI, United States

This panel discussion will focus on careers outside of engineering and business schools at the intersection of IEOR and Healthcare. The panel will feature both academic faculty, from schools of public health and medical schools, and industry professionals who have obtained their terminal degree in IE/OR/MS, and will showcase their career experience. It will also feature advice for final-year PhDs students who are looking to enter the job market. Each panelist will introduce themselves and talk briefly about their career trajectory before opening the panel to questions from the audience.

5 - Panelist

Lance Waller, Emory University, Atlanta, GA, United States

ME42

Building B Level 3 B302

Advancing Cancer Care with Data Driven Approach

Invited Session

Health Applications Society

Chair: Sayaka Ishizawa, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, United States

1 - Validation of Integrating Non-Smoking Risk Factors into Engage Using the NLST Dataset

Maryam Eghbalizarch, Kennesaw State University, Marietta, GA, United States, C. Martin Tammemagi, Sayaka Ishizawa, Seyyed Mostafa Mousavi Janbeh Sarayi, Rafael Meza, Iakovos Toumazis

The ENGAGE framework uses a partially observable Markov decision process to optimize lung cancer (LC) screening for ever-smokers at a personalized level by balancing the potential benefits and harms of screenings. A limitation of ENGAGE is that it only considers age, sex, smoking status, smoking intensity, and years since smoking cessation for dynamically estimating LC risk over time. This study extends the original ENGAGE framework by integrating non-smoking related risk factors (i.e. race/ethnicity, body mass index, education, family and personal cancer history, COPD, and comorbidities including diabetes, hypertension, stroke, and cardiovascular disease) into its risk estimation model. Three versions of Cox regression models (6-race, 3-race, and no-race PLCOENGAGE2025) were developed using data from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) to predict LC incidence, and one model to predict death from non-LC causes. These models were validated both on the PLCO intervention arm and the National Lung Screening Trial (NLST). We integrated these risk models into ENGAGE and derived personalized screening schedules for NLST subjects and compared the effectiveness and efficiency of ENGAGE versus NLST. ENGAGE using the Bach model reduced the number of screening exams to 81% and false positives to 84%, while delaying 2.8% of screen-detected LC diagnoses compared to the NLST. ENGAGE-derived schedules based on the PLCOENGAGE2025 risk models reduced screening exams to 79.6%–80% and false positives to 82.1%–82.5%, while delaying 2.7%–3.3% of screen-detected LC diagnoses. Therefore, including non-smoking risk factors in the risk models can enhance effectiveness of screening across diverse cohorts.

2 - Will Machines Take over? Algorithms for Human-Machine Collaborative Decision Making in Healthcare

Mehmet Ahsen, University of Illinois at Urbana-Champaign, Champaign, IL, United States, Mehmet Ayvaci, Radha Mookerjee

This study presents a cost-effective diagnostic system for allocating mammography interpretation tasks between AI and radiologists, ensuring patient

safety. Using a two-threshold policy based on algorithmic risk scores, it minimizes costs by combining automation and delegation strategies. Back-testing shows potential savings of 20.9%, with broader implications for AI-human collaboration workflows.

3 - Metamodeling Population Health Outcomes: Accounting for Varied Crc Screening Behaviors and Intervention Strategies

Ashley Stanfield, North Carolina State University, Raleigh, NC, United States, Maria Mayorga, Meghan O'Leary, Kristen Hassmiller Lich

Colorectal cancer (CRC) is the third leading cause of cancer-related deaths in the United States. When detected early through screening, CRC can often be treated effectively or prevented by removing precancerous lesions during a colonoscopy. Routine screening—either with fecal immunochemical tests (FIT) or colonoscopy—is recommended for individuals aged 45 to 74. However, as of 2023, only 58% of age-eligible individuals were up-to-date with screening. Increasing adherence is a public health priority, as higher screening rates are associated with more cancer deaths averted. A validated simulation model has been used to estimate health outcomes under different screening scenarios. Using this model, we previously developed metamodels to rapidly estimate outcomes such as cancer cases and life years lost for individuals who originally screened with FIT and received a FIT-based intervention. In this work, we extend the framework to account for population subgroups defined by three factors: their original screening modality (FIT or colonoscopy), the type of intervention they receive (FIT or colonoscopy), and their screening level. We estimate outcomes for each subgroup and then aggregate these to generate population-level results. This enables comparisons between strategies such as targeting all individuals with FIT, all with colonoscopy, or tailoring the intervention to each subgroup to maximize health gains. We also identify the screening increases required across intervention-modality combinations to achieve similar outcomes. This approach helps decision-makers evaluate and compare targeted interventions based on real-world screening behaviors.

4 - Machine Learning-Based Models for Personalized Decision Making in Acute Myeloid Leukemia

George Atherton, The University of Oklahoma, Norman, OK, United States

Acute myeloid leukemia (AML) remains the most common type of acute leukemia cases in adults and affects more than 21,000 people in the US each year with incidence rates projected to rise. Existing treatment protocols prescribe induction chemotherapy (and possibly radiation therapy) with the aim of achieving complete remission. However, certain patients may require hematopoietic stem cell transplantation. Despite its desirable overall survival outcomes in the AML patients, the presence of comorbidities can complicate the transplant leading to negative outcomes, particularly non-relapse mortality (NRM). In this talk, we will explore how machine learning models can support personalized treatment decisions for AML by predicting NRM risk. Specifically, we utilize data from AML patients treated at the OU Stephenson Cancer Center to train and evaluate our proposed models.

ME43

Building B Level 3 B303

AI Models in Healthcare

Invited Session

MSOM: Healthcare

Chair: Hossein Piri, Haskayen School of Business-University of Calgary, Calgary, AB, Canada

1 - Navigating Spatial Queueing Networks with Travel and Delay Tradeoffs

Marco Bijvank, University of Calgary, Calgary, AB, Canada, Sandjai Bhulai

We study a queueing network where customers decide which physical queue to join based on announced delay information and their respective travel times to each queue. The waiting times of each queue are approximated by modeling the system as a Markov-modulated Poisson process. Additionally, each queue location can decide to inflate or deflate their announced wait time from the predicted wait time. However, queue locations are getting a reward based on customer satisfaction that is determined by the over- or underperformance compared to the announced wait time. As a result, servers have to make a trade-off between customer throughput and customer satisfaction. An obvious application is a network of Emergency Departments (EDs) where patients can decide which ED location to visit.

2 - Optimizing Ambulance Location and Relocation in Conflict Zones: a Case Study from Northwest Syria

Sara Ahmed, University of Toronto, Toronto, ON, Canada, Sheng Liu, Timothy Chan

In conflict-affected regions, ambulance services face unique challenges. Unlike typical emergency systems, operations in these settings are decentralized, with multiple stakeholders managing separate fleets, each with different priorities. Ambulance services must operate under sudden demand surges triggered by conflict, along with fluctuating vehicle availability. This study considers ambulance planning over a multi-period horizon, motivated by operations in Northwest Syria, where ambulances respond to both emergency calls and inter-facility transfers. These two service types differ in dispatch rules, operational requirements, and how their demand varies with conflict dynamics. The first part of the work focuses on empirically modeling how changes in conflict intensity affect service demand across space and time, using historical conflict and ambulance data. Building on this, a decision model is developed to determine ambulance placements and relocations, accounting for conflict-driven shocks and partner-specific objectives. The study is part of ongoing work with the World Health Organization and implementing partners, aiming to support more adaptive, data-driven planning in decentralized, high-risk settings.

3 - Behavior-Based Price Discrimination in AI Product Markets: Incentivizing User Contributions Under Duopoly Competition

Bo Feng, Soochow University, Suzhou, China, People's Republic of, JIXIN ZHAO, Tianhu Deng

In user-generated markets, such as those for AI products and databases, product quality evolves with user participation and effort. This paper studies how firms can leverage these user driven quality improvements using behavior-based price discrimination (BPD) in a two-period vertical duopoly comprising a private firm and an open-source firm. In Period 1, users choose between purchasing the private firm's established product or by adopting the open source firm's base product and contributing effort to improve its future quality. Before Period 2, the open-source firm announces a contributor discount; both firms then set prices that discriminate based on user purchase

histories. We find that higher contributor discounts do not always incentivize greater user effort, as strategic users may anticipate future price reductions by the private firm and delay contributions. The optimal discount for the open-source firm increases with the marginal quality gain from user effort but decreases with greater user heterogeneity or a larger incumbent quality lead. Contribution-based BPD improves consumer surplus and total welfare, but can erode firm profits if discounts are misaligned. These results highlight key strategic trade-offs: open-source firms should link discounts to the quality value each user generates and avoid overly aggressive incentives that provoke competitive undercutting. Private firms can counter by targeting poaching strategies, while policymakers aiming to promote innovation and welfare may support stronger contributor rewards through subsidies.

4 - Mitigating Missing Outcome Bias in Multi-Site Healthcare Data Using Auxiliary Outcomes for Wait TIME Estimation

Jae Park, University of Central Florida, Orlando, FL, United States, Antonio Castellanos, Joseph Zourob

Building machine learning models using multi-site healthcare data is essential for ensuring generalizability. Models trained on data from a single site may be limited, as their findings might not be applicable to other institutions. To enhance generalizability in healthcare research, federal agencies (e.g., the Agency for Healthcare Research and Quality) manage data from numerous healthcare institutions across the United States. Additionally, data consortiums (e.g., OneFlorida) establish data-sharing agreements with multiple institutions, centralizing data collection from affiliated sites.

Despite these efforts, multi-site data frequently suffer from quality issues due to inconsistencies in clinical practices and data availability. For example, some hospitals collect a particular health outcome while others do not, leading to missing data. This can result in biased models that perform better for hospitals with more complete data. In this study, we tackle this challenge by leveraging an auxiliary outcome that is closely related to the primary outcome of interest.

Our numerical experiments focus on estimating emergency department (ED) waiting times using the national ED data from 2016 to 2022. Among the 1,219 hospitals in the dataset, 57 do not report waiting time data, and 9,464 (9.5%) out of 99,491 records are missing this outcome. Since the length of ED visit provides an upper bound on the wait time, it can serve as an auxiliary outcome. By incorporating this additional information, we enhance prediction performance across participating hospitals and mitigate the bias introduced by incomplete data.

5 - Detecting Positive Cognitive Aging Using Diffusion MRI

Guiping Hu, Oklahoma State University, Stillwater, OK, United States, Mohammad Fili, Parvin Mohammadiarveje, Auriel willette

While most adults experience some degree of age-related cognitive decline, a subset known as “Positive-Agers” show superior cognitive abilities compared to their age-matched peers. Identifying individuals with exceptional cognitive performance versus those experiencing decline is critical for guiding targeted interventions in older populations. In this study, we introduce *Optimal Cognitive Scoring* (OptiCS) algorithm, designed to distinguish Positive-Agers from Cognitive Decliners. Drawing on longitudinal data from 5,797 participants in the UK Biobank, our predictive pipeline achieved strong classification performance, with an area under the curve (AUC) of 0.83. The most informative features were diffusion MRI metrics from tracts associated with pathological aging, including the hippocampal fornix, the tapetum of the corpus callosum, and other key white matter pathways. This work offers three major contributions: (1) a reliable scoring system to detect cognitive shifts, (2) evidence that diffusion MRI can serve as a powerful biomarker for cognitive status in healthy adults, and (3) a predictive framework for the early identification of cognitive decline.

ME44

Building B Level 3 B304

Healthcare Analytics

Invited Session

MSOM: Healthcare

Chair: Guihua Wang, The University of Texas at Dallas, Richardson, TX, United States

Co-Chair: Minmin Zhang, University of Texas at Dallas, Richardson, TX, United States

1 - Impact of Regulation on Deceased-Donor Liver Offer Acceptance and Patient Welfare

Shubham Akshat, University of Houston, Houston, TX, United States, Liye Ma, Christopher Hughes

The public reporting of a transplant program (TP)'s post-transplant outcomes has been a contentious issue in the transplant community in the United States. TPs are also regulated by the Department of Health and Human Services, which periodically monitors their post-transplant outcomes. While they aim to uplift the underperforming TPs and enable informed decision-making (by patients and insurance companies), there can be important unintended consequences. To maintain a good public image and comply with regulations, a TP might deviate from fulfilling its core responsibility of maximizing patient welfare when deciding to accept an organ offer for a patient.

We develop a novel empirical framework to study the impact of regulations on deceased-donor liver offer acceptance decisions at a TP. Our framework integrates patient-centric and regulatory considerations, whereas existing literature has focused only on the former. Further, we incorporate a TP's response to its regulatory standing into a discrete-event simulation to evaluate the impact of regulations on organ non-utilization and patient welfare.

We find that a TP becomes increasingly selective in its offer acceptance decisions as its regulatory standing worsens. This behavior exhibits a near-monotonic pattern for marginal quality organs. Broadly speaking, a TP deviates from a patient-optimal decision more when an organ is of marginal quality. When a TP was in poor regulatory standing, the non-utilization rate of marginal quality organs more than doubled (w.r.t. intermediate standing). More patients died on the waiting list (40 more deaths annually), disproportionately affecting high-risk patients. Lastly, we outline several policy implications.

2 - Causal Evidence of Task-Switching Cost in Organ Transplantation

Yiwen Jin, Haskayne School of Business, University of Calgary, Calgary, AB, Canada, Jiayi Liu, Joel Adler

Laboratory studies have demonstrated that switching between tasks increases cognitive load and error rates, yet its consequences in high-stakes professional contexts remain unclear. Here, we provide causal evidence of the impact of task switching in organ transplantation. Leveraging the unpredictable arrivals of donor organs as a natural experiment, we analyze national data on 316,743 abdominal transplants performed by 982 surgeons over 13 years. We find that switching between organ types (e.g., from liver to kidney) significantly increases one-year post-transplant mortality by 0.58 percentage points (a 12.3% relative increase). These adverse effects can be substantially reduced through structured scheduling practices, increased surgeon experience, and longer intervals between surgeries. Our findings identify task switching as a critical, previously unrecognized risk in organ transplantation and highlight the need for training programs and clinical practices to implement targeted interventions to improve patient survival.

3 - A Heterogeneous Transition Forest Approach for Analyzing Changes in Health States: Implications for Organ Transplantation

Minmin Zhang, The University of Texas at Dallas, Richardson, TX, United States, Guihua Wang

State transitions are crucial for modeling changes in conditions such as patient health. While existing models use this concept in decision-making, they typically assume homogeneity across all units, overlooking individual variability—particularly important in contexts like disease progression. This study introduces a novel method, the Heterogeneous Transition Forest (HT forest), to capture patient-specific transitions in health states. We apply this approach in the context of organ transplantation, modeling how patients' health evolves while waiting for a transplant. Our findings reveal significant variability in transition patterns based on individual patient features, underscoring the limitations of homogeneous models. To evaluate the impact of this heterogeneity, we develop a Markov decision process that incorporates personalized transition probabilities. The results show that using patient-specific transitions in decision-making significantly improves patient outcomes, measured in total expected life days. This research offers a new methodological contribution for estimating heterogeneous transition probabilities, which has practical implications for healthcare. By incorporating individual patient characteristics into decision models, healthcare providers can enhance the effectiveness of treatment strategies and potentially extend patient lifespans.

4 - Data-Driven Approach to the Surgical Block Auto-Release Decision

Sandeep Rath, University of North Carolina at Chapel Hill - Kenan Flagler, Chapel Hill, NC, United States, Moe Lim

Surgical block allocation balances surgeon scheduling and OR utilization. Unused blocks are auto-released for open access before surgery, requiring a trade-off between predictability and efficiency. Using historical booking data, we developed a data-driven approach to optimize auto-release timing, enhancing OR utilization without compromising scheduling predictability in a large medical center

ME45

Building B Level 3 B305

Socially Responsible Service Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Dawson Kaaua, Georgetown University, Washington, DC, United States

Co-Chair: Vanitha Virudachalam, University of Illinois Urbana-Champaign, Champaign, IL, United States

1 - Dining Disrupted: the Impact of Food Delivery Platforms on the Restaurant Industry Market Structure

Sandeep Srinivas, University of Illinois at Urbana Champaign, Champaign, IL, United States, Ujjal Mukherjee, Gautam Pant

Food delivery platforms have transformed the restaurant industry and consumer dining habits. This study examines how their introduction has reshaped the restaurant industry's market structure. Specifically, we focus on which types of restaurants—full-service restaurants versus others—stand to gain or lose from the presence of these platforms. Using multiple empirical techniques, we show that platform entry has shifted market share away from non-full-service restaurants and toward full-service establishments. To uncover the mechanisms behind this shift, we analyze Google reviews and find that non-full-service restaurants receive systematically lower ratings after platforms arrive. We examine different aspects of consumer reviews to understand specific attributes driving this change. Our findings clarify the dominant mechanisms at work in the food delivery platform context and offer strategic insights for restaurateurs and investors to navigate the expanding food delivery market.

2 - Estimating Scope 3 Emissions via Supply Chain Tracing

Sanjith Gopalakrishnan, McGill University, Montreal, QC, Canada, Michael E. Raynor

In the pursuit of global net-zero emissions, hard-to-abate (HTA) upstream sectors like steel and cement pose significant challenges. Meanwhile, several consumer-facing firms in sectors such as retail and consulting have adopted ambitious net-zero supply chain targets. However, these efforts often rely on ad hoc supply chain tracing that rarely extends to deep upstream tiers, thereby excluding high-emission HTA sectors and undercounting supply chain emissions, i.e., Scope 3 emissions. This work therefore addresses a critical challenge in environmental supply chain management: the accurate estimation of Scope 3 emissions. We describe a hybrid carbon footprinting method that combines conventional estimation methods that rely on supply chain tracing with an input-output life cycle analysis approach. By situating a traditional supply network model within an input-output framework, we map the problem of identifying the optimal supply chain tracing strategy to a graph-theoretic optimization problem and provide an exact solution approach to improve the accuracy of Scope 3 emissions estimates.

3 - A Fast Ecological Inference Algorithm for the R times C Case

Charles Thraves, University of Chile, Santiago, Chile, Pablo Ubilla, Daniel Hermosilla

The RxC ecological inference problem is known to be difficult to address. In the past, researchers have approached this problem from various angles, including parametric probability models, entropy-maximization, and mathematical programming. In this work, the ecological inference problem is based on an election context where, at each ballot box, we observe candidates' votes and the number of voters from each demographic group. We estimate non-parametric voting probabilities for each group by applying the EM algorithm, in which the M-Step can be solved in closed-form, whereas the E-Step can only be solved in practice for small instances. We show that the M-Step can be solved with a closed-form solution, while the E-Step requires an exponential number of steps to be solved exactly. To address this, we propose four approximation methods to compute the E-Step in polynomial time. Through simulated instances, we observe that the resulting probability estimates using these methods are very close to the real values, with running times less than a thousandth of a second. We then introduce a methodology to perform group aggregation when the observed data is insufficient to estimate voting probabilities accurately. We apply this technique to the 2021 Chilean Presidential election, obtaining estimates of voting probabilities for different age ranges. We observe that older voters tend to support right-wing candidates and long-established parties, whereas younger voters show the opposite tendency. The described method is also used to detect outlier ballot boxes. Developed methods are available in the R library 'fastei'.

4 - Coproduction in the Classroom: Optimally Allocating Incentives Between Teachers and Students

Vanitha Virudachalam, Gies College of Business, UIUC, Champaign, IL, United States, Nan Liu, Sergei Savin

Schools have long experimented with monetary incentives to encourage high levels of teacher effort. Less common is the use of monetary incentives for students, even though education is a coproduced service in which both teachers and students must exert effort to achieve a successful outcome. We study the problem faced by a school district that must optimally allocate a monetary incentive between teachers and students to achieve a baseline threshold of student proficiency levels. In our model, teachers manage a classroom of N students, where there are two types of students: high- and low-performing students. Teachers must choose their optimal effort allocation between the two types of students, whereas students individually determine how much effort to exert. For both teachers and students, effort is costly. Using a single-period Bertrand model, we study the optimal allocation of a merit-based incentive between teachers and students and the resulting effort decisions of each party.

5 - A Moment for Reflection: De-Biasing Service Evaluations

Hallie Cho, Georgetown University, Washington, DC, United States, Dawson Kaaua

Problem definition: Service evaluations typically solicit feedback in two ways: star ratings and written feedback (i.e., comments). Although star ratings are typically collected before comments, we propose that switching this order to ask for comments first can mitigate bias present in star ratings since the process of writing provides time and space for the evaluator to reflect on their experience. Methodology/results: We conducted an experiment on Amazon Mechanical Turk where 4261 participants received a sequence-based service, with the same overall level of service, from servers who varied in their demographic characteristics. Following the service, the participants were asked to evaluate the performance of the servers with the order of star ratings and written feedback randomized. We find no evidence of demographic bias toward the servers (e.g., race, gender, sexuality, age, and body size) but find that the sequence of good and bad experiences in the service leads to biased star ratings. Most importantly, we find that collecting comments prior to star ratings effectively mitigates the detected sequential bias in subsequent star ratings due to participants reflecting on the service experience. Managerial implications: Our paper shows that the most widely used quantitative feedback (i.e., average star ratings) can be biased if collected in the traditional manner and that this bias can be reduced if qualitative feedback (i.e., comments) is collected first. Implementing such a change in the way organizations conduct service evaluations could yield many benefits.

ME46

Building B Level 3 B306

Decision Analysis Award session

Award Session

Decision Analysis Society

Chair: Robin Dillon-Merrill, Georgetown University, Washington, DC, United States

1 - Awards Session**Robin Dillon-Merrill, Georgetown University, Washington, DC, United States**

Award Session Chair

2 - Student Paper Award Chair**Rupert Freeman, University of Virginia, Charlottesville, VA, United States**

Award Chair

3 - Publication Award Chair**alessandra cillo, LIUC University, Castellanza (Va), Italy, Johannes Muller-Trede**

The Decision Analysis Society Publication Award is given annually to the best decision analysis journal article or book published in the second preceding calendar year, as judged by an award committee. Publications appearing in the year 2023 would be eligible for consideration in the year 2025. For this award, decision analysis is defined as a prescriptive approach to provide insight for decision making based on axioms that are logically consistent with the axioms of von Neumann and Morgenstern and of Savage. Key constructs of decision analysis are utility to quantify one's risk preferences and probability to quantify the state of one's knowledge. The intent of the award is to recognize the best publication in decision analysis. Contributions could include, but are not limited to, theoretical, methodological, and procedural contributions to decision analysis, descriptions of applications and experimental studies. Nominated publications will be judged for significance, relevance, originality, and readability.

4 - Ramsey Medal Award Chair**Yael Grushka-Cockayne, Darden School of Business, University of Virginia, Charlottesville VA, VA, United States**

The Frank P. Ramsey medal is awarded for distinguished contributions in decision analysis. Distinguished contributions can be internal, such as theoretical or procedural advances in decision analysis, or external, such as developing or spreading decision analysis in new fields.

5 - Das Treasurer**Onesun Steve Yoo, University College London, London, United Kingdom**

?

ME47

Building B Level 3 B307

Editorial Insights: A Panel Editorial Insights: A Panel Discussion with MSOM/Healthcare Journal Editors

Panel Session

Manufacturing and Service Operations Management (MSOM)

Co-Chair: Yichuan Ding, McGill University, Montreal, QC, Canada

Co-Chair: Susan Lu, University of Toronto, Toronto, ON, Canada

1 - Moderator Panelist**Susan Lu, University of Toronto, Toronto, ON, Canada****2 - Moderator Panelist****Yichuan Ding, McGill University, Montreal, QC, Canada****3 - Panelist****Andrew Schaefer, Rice University, Houston, TX, United States****4 - Panelist****Lauren Lu, Dartmouth College, Hanover, NH, United States****5 - Panelist****Ozlem Ergun, Northeastern University, Newton, MA, United States****6 - Panelist****Sarvanan Devaraj, University of Notre Dame, Notre Dame, IN, United States****ME48**

Building B Level 3 B308

Frontiers of Information Design, Regulation Design, and Decision Making

Invited Session

Decision Analysis Society

Chair: Qian Zhang, California State Polytechnic University Pomona, Pomona, CA, 91767, United States

1 - Optimal Regulation Paradigm Design with Privacy Protection and R&D Incentive**Qian Zhang, California State Polytechnic University Pomona, Pomona, CA, United States, Luying Wang**

Consumer privacy refers to the safeguarding of personal information shared by individuals with businesses, ensuring that such data is not misused or improperly disclosed. This has led to public discussions on whether individuals' privacy should be better protected by restricting the amount or the precision of information that commercial institutions collect on their participants. We design a model by addressing the impact of optimal privacy regulation on the firm's production investment in the context of Bayesian persuasion, where one regulator can design an information revelation policy to reveal a certain degree of consumers' private information. A monopolistic firm observes the signal about consumers, then makes a recommendation and price offer to uninformed consumers who have recourse to some outside option. Since such a signal reveals certain private information about consumers, and the firm cannot learn more than the signal, we define such signal as privacy regulation, i.e., the maximum amount of consumer information that can be revealed to the firm. We show that if the regulator's objective is to maximize a combination of consumer surplus and R&D investment, then there always exists a unique optimal privacy protection policy to reveal a certain degree of consumer privacy information to the firm. Further, if the

regulator is able to further design other regulations such as minimum quality regulation or price fairness regulation, then potentially it is possible to further improve social welfare. We show that it is optimal for the regulator to design a combination of regulations endogenously to maximize social welfare.

2 - Optimal Data Pricing Decisions of Competitive Two-Sided Platforms with Heterogeneous Data Costs

Jinpu Li, Tsinghua University, Beijing, China, People's Republic of, Ronghua Sui, Ruize Gao

Two-sided platforms play the role of data intermediaries in the era of the digital economy. Data are byproducts of users' activities and are replicable, creating an important profit source and competitive advantage for the platforms. In this paper, we build a Hotelling model to study data collection and the related pricing of competitive platforms with heterogeneous costs. Our study identifies differences in the pricing scheme based on whether platforms adopt consistent or differentiated collection strategies, and examines the impact of data usage and affiliation behavior on platform decisions. Our main findings are as follows: First, when one platform holds a significant cost advantage, adopting differentiated data collection strategies for both platforms is optimal; otherwise, consistent strategies are preferable. Notably, in scenarios where both platforms adopt consistent collection decisions (i.e., both partially collect data), the cost-disadvantaged platform benefits more from increased data collection compared to its cost-efficient counterpart, heavily influencing its pricing based on internal costs. Second, for any platform, lowering pricing can become a wise strategy as the value derived from data usage increases under certain condition. This approach not only attracts more users but also enhances overall profitability by leveraging the higher utility that each unit of data provides. Third, two-sided users' affiliation behaviors will influence the platforms' data decisions in distinct aspects. The change of consumer affiliation behavior primarily impacts the platforms' data pricing, whereas provider affiliation behavior will change data collection strategies significantly.

3 - The Design and Price of Influence

Raphael Boleslavsky, Indiana University, Bloomington, IN, United States, Aaron Kolb

A sender with private preferences would like to influence a receiver's action by providing information through a statistical test. The technology for information production is controlled by a monopolist intermediary, who offers a menu of tests and prices to screen the sender's type. We characterize the intermediary's optimal screening menu and the associated distortions, which may benefit the receiver by increasing test informativeness. Because of these distortions, seemingly unfavorable changes in the prior belief may actually benefit the sender. In extensions, we study (i) a stronger intermediary who can commit to a "coercive" test to punish non-participation and (ii) a weaker intermediary who cannot control test design but can charge for access.

4 - Your Data, My Data: Information Disclosure and Competition in Marketplace Platforms

Noam Shamir, Tel Aviv University, Tel Aviv, Israel, Sarit Markovich, Yaron Yehezkel

The dominance of e-commerce platforms has raised concerns about potential anti-competitive practices, particularly due to their dual roles as marketplace operators and sellers. This study examines when sellers choose to disclose their private information concerning characteristics of potential buyers to a market place platform. Such disclosures can result in "informational spillovers," where platforms exploit this information to compete against sellers. This dynamic may discourage sellers from sharing critical information, leading to market inefficiencies. We analyze the conditions under which these inefficiencies arise and evaluate two policy remedies: vertical separation and informational firewalls. Our findings reveal that sellers choose to conceal buyer information under moderate substitution levels between products to prevent strategic competition, resulting in welfare losses due to reduced market competition and inefficient advertising allocation. Vertical separation, while encouraging information sharing, can reduce competition, resulting in mixed welfare outcomes. Informational firewalls, by restricting platforms from using seller information to gain a competitive advantage, promote consistent information sharing but may, in some cases, also limit competition and lead to inefficient advertising allocation. These findings highlight the intricate interplay between platform competition, seller information sharing, and regulatory interventions, offering valuable insights for policymakers and businesses operating in platform-driven markets. Our study also underscores the tension between maximizing welfare and protecting seller's private information because welfare is highest when the platform utilizes the seller's information. We also show that in some cases, the seller prefers the platform to use its private information because doing so leads the platform to avoid competition.

ME49

Building B Level 3 B309

Innovative Strategies in Supply Chain and E-Commerce Operations

Invited Session

MSOM: Supply Chain

Chair: Lu Kong, University of South Florida, Tampa, FL, United States

Co-Chair: Ting Wang, University of Science and Technology of China, Hefei, N/A

Co-Chair: Kejia Hu, University of Oxford Said Business School, Oxford, United Kingdom

1 - Competitive Partial-Payment Presale with Consumer Switching

xuhan tian, Shanghai Tech University, Shanghai, China, People's Republic of, Xiangtong Qi, Yusen Xia

Common E-commerce platforms have increasingly adopted partial-payment presale promotions as a marketing tactic. This promotion allows consumers to secure a product by paying a non-refundable deposit before the product is released, capitalizing on consumers' uncertain and often fluctuating purchasing intentions. The inconsistency of consumers' purchasing intentions is reported to be amplified during highly competitive sales events, manifesting as impulsively pre-ordering after being exposed to targeted pop-up ads, frequent regretting and switching to later found alternatives. Such behaviors have added a new structural element to the conventional presale competition, impacting online sellers' profitability and calling for a focused study for the platform to enhance promotion design and management. We construct an analytical model to study the presale competition in the presence of switching consumers, from which we further investigate how sellers' profitability may be affected by how the presale is organized. Our results suggest that competition can drive sellers to prevent consumers from switching through price reductions, potentially diminishing presale advantage and reducing profitability. However, the platform can mitigate this effect by strategically facilitating consumer switching through appropriate deposit control, which can also increase social welfare with higher equilibrium profits and consumer

2 - Collaborative Prepositioning with Postponement and Stock Sharing in a Network with Regional and Country Warehouses

Lamia Gulnur Kasap Simsek, NA, Dallas, TX, United States, Birce Adsanver, Burcu Balcik, Marie-Ève Rancourt

We evaluate the impacts of postponement and stock-sharing strategies among humanitarian organizations (HOs) operating within regional and country-level depots for prepositioning supplies. In our setting, HOs in both regional and country warehouses respond to disasters within their designated areas in a specific region. Our proposed system involves HOs stocking supplies in both branded (final product) and unbranded (shareable standard) forms at the regional depot, while limiting stocks to branded form only at country depots for faster response. To operationalize postponement and stock sharing, we develop a two-stage stochastic optimization model that minimizes expected response time weighted by disaster severity under fixed total base stock constraints. This model optimizes the levels of branded and unbranded items for each HO across warehouses, facilitates stock sharing in the regional warehouse, and devises optimal stock mobilization strategies to affected countries. We test our approach using historical hurricane scenarios from the Caribbean region and data from HOs. We demonstrate substantial savings in fill rate, response time, and inventory utilization.

3 - Serving E-Commerce and E-Grocery Deliveries Together with Learning-Based Vehicle Dispatching

Oguz Sohret, Koc University, Istanbul, Turkey, Baris Yildiz, Gurhan Kok

We present a machine-learning based approach to support routing optimization in a collaboration model between a grocery and an e-commerce company, both of which agree to deliver online orders together using a single vehicle. Grocery orders including same-day delivery are delivered within time windows, whereas e-commerce orders are delivered anytime in the day.

ME50

Building B Level 3 B310

OM in Education

Invited Session

MSOM: Service Operations

Chair: Helen Wang, London Business School, London, United Kingdom

1 - Mitigating Gender Bias in Higher Ed Instructor Evaluations

Tong Wang, Chicago Booth, Chicago, IL, United States, Monika Heller, Kamalini Ramdas

Instructor ratings play a crucial role in learners' course selection and significantly impact key career outcomes for instructors, including recruitment and promotion. While gender bias in traditional classroom evaluations is well-documented, operational strategies to mitigate such bias have received limited attention. We ran three preregistered online experiments to examine how instructor gender and prior ratings impact instructor evaluations, under identical teaching quality. Our findings show that both female and male learners exhibit bias against instructors of the opposite gender, mainly when they face a higher cognitive load due to time pressure. Furthermore, male learners tend to correct unfairly low ratings only for male instructors, introducing a gender bias against female instructors. Learners exposed to high prior ratings tend to increase their ratings for instructors of the same gender, exhibiting confirmation bias. Our findings that publishing prior ratings or increasing cognitive load could exacerbate gender inequity highlight two previously unexplored operational levers to mitigate gender bias.

2 - The Impact of Race-Blind and Test-Optional Admissions on Racial Diversity and Merit

Allen Sirolly, Columbia Business School, New York, NY, United States, Hongyao Ma, Yash Kanoria

How significant was the role of racial preferences in U.S. college admissions before the Supreme Court's 2023 decision to ban race-based affirmative action? How much might test-optional admission policies impact racial diversity and academic merit? In this work, we estimate a simple model of college admissions decisions from 2012–2020, leveraging a novel dataset of applicant profiles and admissions outcomes across the full spectrum of college selectivity. We find that, broadly, the impact of race and testing policies on diversity and merit of admits decreases by college selectivity. For America's less selective colleges that collectively enroll over three-quarters of students, fully eliminating preferences for race (and for unobserved factors which are correlated with race) has little impact on the proportion of underrepresented minorities (URM) and on the average SAT score of admitted students. In contrast, for the 34 most selective colleges accounting for 3 percent of total enrollment, our estimates suggest that admissions going "race blind"—absent any compensating changes in admissions criteria—could reduce URM admission by one-third while increasing the average SAT score of admits by no more than 10 points. We also estimate that universal test-optional admission may lead to a small increase in the proportion of URMs at the most selective colleges, while decreasing the average SAT score by up to 10 points. At less selective institutions, the effects are estimated to be negligible.

3 - Non-Profit Support in Education: Resource Allocation and Students' Lifetime Outcomes

Harish Guda, Arizona State University, Tempe, AZ, United States, Goutham Takasi, Milind Dawande, GANESH JANAKIRAMAN

Our work focuses on the operations of non-profit organizations (NPOs) that broaden access to high-quality education for underprivileged students. Specifically, we analyze the resource-allocation strategy of an NPO that adopts a two-stage structure in allocating resources to its beneficiaries, e.g., free pre-secondary education (first stage) for all students in a target population, followed by sponsorships for post-secondary education (second stage) at leading institutions to students who demonstrate commendable performance in the first stage. The lifetime outcomes of the beneficiaries depends on their own effort and on the quality of the resources that the NPO provides.

We establish the strategic role of an NPO's resource-allocation strategy on the effort beneficiaries exert and their lifetime outcomes. In particular, despite possessing an adequate quantity of resources to support all its beneficiaries, we show why the NPO benefits from *strategically throttling access* to the resources. Our findings have implications for the design of such support policies of NPOs. For a fixed endowment of resources, we demonstrate the effect of competition among the beneficiaries on their effort and lifetime outcomes, and show the value of creating a *strategic scarcity* of resources for incentivizing beneficiaries to exert more effort. Finally, when faced with multiple beneficiary subgroups, we identify when the NPO benefits from pooling the beneficiary subgroups vs. earmarking dedicated resources for each subgroup.

4 - Teaching at a Distance, Scrutinized up Close: Bias in the Online Student Evaluation of Teaching

Lu Kong, University of South Florida, Tampa, FL, United States, Kejia Hu, Lanfei Shi

This study investigates how the shift to online education has influenced student evaluations of teaching (SETs), with a focus on whether instructor gender and racial biases have intensified compared to in-person settings. SETs remain a central tool for assessing faculty performance in U.S. higher education, yet they are prone to systematic bias. Understanding how digital learning environments shape these biases is critical for ensuring fair and effective evaluation systems. Drawing on 598,252 teaching evaluations across the top 150 U.S. colleges, we employ a Difference-in-Differences-in-Differences (DDD) model with Propensity Score Weighting to examine changes in evaluation scores before and after the COVID-19-induced transition to online instruction. We supplement this analysis with text mining of student comments to uncover behavioral mechanisms driving disparities. Our results reveal that non-white male instructors experienced a significant and persistent decline in ratings—0.69 points on a 5-point scale, or 16.2% lower than white male counterparts—following the move to online teaching. This disparity was most pronounced in institutions with low student and faculty diversity, in comments containing racial or gender-biased language, and among students who valued entertainment over educational substance. These findings suggest that online education can exacerbate existing biases by increasing psychological distance and reducing interpersonal cues. To mitigate these effects, we propose institutional strategies including the adoption of multi-method evaluation frameworks, redesign of SET instruments to raise bias awareness, and a stronger emphasis on inclusive hiring and curricular practices. Our study offers actionable insights for promoting fairness in digitally mediated teaching evaluations.

5 - Enhancing Student Engagement Through Gamified Reward Consumption Design: A Field Experiment in Education

Xiao Lei, The University of Hong Kong, Hong Kong, Hong Kong, Yimin Yin, Yipu Deng

We introduced a points-redemption feature into a school system that had previously only recorded the points students earned for academic tasks. To test its impact, we ran a four-month field experiment in four matched primary-school campuses. Students earned points for completing homework and correcting errors. Campuses were randomly assigned to one of four conditions: control (points only recorded), permanent purchase (points buy virtual cards kept forever), temporary purchase (cards expire after two weeks), or random draw (points buy a lottery card). Tracking weekly homework completion, accuracy, and error correction shows that gamification boosts performance overall, but the effects differ across conditions, underscoring the importance of game-element design in educational gamification.

ME51

Building B Level 3 B311

Market Design, Platforms, and Matching Markets

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Francisco Castro, UCLA Anderson School of Management, Los Angeles, CA, United States

Co-Chair: Alfredo Torrico, Cornell University, New York, NY, United States

Co-Chair: Alfredo Torrico

1 - Optimal Sequential Screening with Private Budgets

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

We study the problem faced by a seller of designing an optimal sequential screening mechanism when a buyer's budget is privately known. Our mechanisms allow for sequential screening in that the buyer first reports their budget and, conditional on their budget, the mechanism determines the allocation and payment rule. For deterministic mechanisms, we show that each budget type can be associated with a fractional quantity of the item. From this, we derive a tight upper bound on the seller's profit by reformulating the problem as choosing a concave price function of the quantity purchased by each type. We further simplify the problem by proving that this optimal price function must be linear and characterized by the solution to a one-dimensional optimization problem. Consequently, the optimal mechanism is a pooling contract. For randomized mechanisms, using a similar approach to the deterministic case, we find that the optimal mechanism is also a pooling contract, although it may incorporate randomization for some valuation ranges.

2 - Revenue Maximization in Choice-Based Matching Markets

Alfredo Torrico, Cornell University, New York, NY, United States, Dan Nissim, Danny Segev

The primary contribution of this paper resides in devising constant-factor approximation guarantees for revenue maximization in two-sided matching markets, under general pairwise rewards. A major distinction between our work and state-of-the-art results in two-sided assortment optimization is that, for the first time, we are able to address *reward maximization*, reflected by assigning each customer-supplier pair an arbitrarily-valued reward. The specific type of performance guarantees we attain depends on whether one considers the *customized* model or the *inclusive* model. The fundamental difference between these settings lies in whether the platform should display to each supplier all selecting customers, as in the inclusive model, or whether the platform can further personalize this set, as in the customized model. Technically speaking, our algorithmic approach and its analysis revolve around presenting novel linear relaxations, leveraging convex stochastic orders, employing approximate dynamic programming, and developing tailor-made analytical ideas. In both models considered, these ingredients allow us to overcome the lack of submodularity, monotonicity and subadditivity that stems from pairwise rewards, plugging the applicability of existing methods.

3 - Personalization vs. Marketplace Diversity: How Recommendation Algorithms Shape Pricing, Participation, and Consumer Search

Anil Omer Saritac, Singapore Management University, Singapore, Singapore, Nur Kaynar, Ozge Sahin

Online marketplaces use recommendation algorithms to steer shoppers toward products, generating about 30 % of page views and sizable welfare gains. Most research optimizes these algorithms for buyer satisfaction, treating sellers as passive and overlooking how recommendations shape pricing, entry, and market diversity. This bias can crowd out niche sellers, erode variety, and weaken long-run platform health.

We build a dynamic search model with a continuum of horizontally differentiated sellers who choose fixed, type-dependent prices to maximize discounted profits, and heterogeneous customers who arrive each period. Upon arrival a customer receives a personalized product list—formally, a probability distribution over sellers—generated by the platform. After clicking, the customer observes match utility and either buys or continues searching, facing a departure probability that captures search costs. The recommendation distribution therefore mediates both seller pricing incentives and customer search paths.

Analytically, we convert the high-dimensional game into a tractable form that reveals clear comparative statics. We characterize how the degree of personalization and an increased emphasis on diversity in recommendations affect market outcomes: Welfare-maximizing or profit-maximizing recommendations do **not** necessarily fully personalize. The first-best balances relevance with deliberate exposure to niche sellers, which sustains marketplace diversity. The model pinpoints when this balance shifts under alternative platform objectives or greater time pressure, and offers a template for other strategic search settings.

4 - Information Design for Spatial Resource Allocation

Ozan Candogan, University of Chicago, Chicago, IL, United States, Manxi Wu

In this paper, we study platforms where resources and jobs are spatially distributed, and resources have the flexibility to strategically move to different locations for better payoffs. The price of the service at each location depends on the number of resources present and the market size, which is modeled as a random state. Our focus is on how the platform can utilize information about the underlying state to influence resource repositioning decisions and ultimately increase commission revenues. We establish that in many practically relevant settings a simple monotone partitional information disclosure policy is optimal. This policy reveals state realizations below a threshold and above a second (higher) threshold, and pools all states in between and maps them to a unique signal realization. We also provide algorithmic approaches for obtaining (near-)optimal information structures that are monotone partitional in general settings.

ME52

Building B Level 3 B312

Sustainable and Socially Responsible Operations

Invited Session

MSOM: Sustainable Operations

Chair: Soudipta Chakraborty, University of Kansas, Lawrence, United States

1 - Buyer Management of a Supplier's Social Responsibility Risk: Prevention, Detection, and Remediation

Junhao Yu, Miami University, Oxford, OH, United States, Tim Kraft

We study how retailers can proactively manage their suppliers' responsibility risk. We focus on a retailer's risk management strategy and consider three tactics the retailer can deploy—prevention, detection, and remediation—and study how combining the three can help the retailer achieve optimal results.

2 - Spreading the Wings: Understanding Optimal Strategies for Social Enterprise Growth

Wei Wei, University of New Hampshire, Durham, NH, United States, Priyank Arora, Gokce Esenduran

We study the growth dilemma faced by a social enterprise that pursues a dual objective: Delivering societal impact for its beneficiaries (typically underserved and vulnerable individuals) while generating profit from paying customers by offering essential services such as healthcare and education. Motivated by the real-world example of Aravind Eye Care System that uses profit generated from paying customers to subsidize service delivery to beneficiaries, we examine the choice between the following two growth strategies: (1) Expanding at the existing location by adding more capacity and offering additional value-adding services to beneficiaries; or (2) Entering a new location by building capacity at a distant location. The former strategy relies on costly investments in adding capacity and expanding the scope of services for beneficiaries, but all the operational elements are known to the enterprise. The latter strategy relies on potentially less costly investment in capacity (due to lower diminishing returns when starting anew), but it also entails uncertainty in the cost of providing services to beneficiaries at the new location (due to localized needs of vulnerable populations). We develop an analytical model to capture these trade-offs and interrelate the social enterprise's investment decisions under each strategy with key operational elements. We identify conditions under which each strategy is optimal. To the best of our knowledge, ours is the first study to provide guidance to social enterprises on whether to grow by enhancing their footprint at the same or a new location, taking into account the operational elements.

3 - Additionality of Carbon Offsets: Project-Specific vs. Standardized Baselines

Soudipta Chakraborty, University of Kansas, Lawrence, KS, United States, Vishal Agrawal, Safak Yucel

A carbon offset represents reduction in emissions that can be used to compensate for emissions elsewhere. Developers generate offsets through emissions-reduction projects. A non-profit carbon registry issues offsets to developers while ensuring additionality: an offset needs to represent one unit of reduction from a developer's business-as-usual emissions, i.e., what the developer's emissions would have been in the absence of the market for offsets. Accordingly, environmental groups raise greenwashing concerns against non-additional offsets. Ensuring additionality is challenging because business-as-usual emissions is a developer's private information and an unobservable counterfactual for the registry. In practice, the registry assigns a baseline to represent business-as-usual emissions through one of the two methods: Under the project-specific method, a developer self-reports its business-as-usual emissions to the registry, which then inspects the report and assigns reported emissions as the baseline if it accepts the project. Under the standardized method, the registry assigns a common baseline to a group of similar developers. We compare economic and environmental implications of these two methods.

ME53

Building B Level 3 B313

Emerging Topics in Revenue Management and Pricing

Invited Session

Revenue Management

Chair: Sajad Modaresi, UNC Chapel Hill, Kenan-Flagler Business School, Durham, NC, United States

1 - Menu Optimization for (Office) Meal Delivery Platforms**Saman Lagzi, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States**, Ningyuan Chen, Pin Gao, Sheng Liu, Chenhao Wang

We study the problem an office meal delivery platform faces every day. Such platforms connect client firms, with catering needs for their employees, to restaurants, by offering a menu of compatible restaurants to all the employees in an office. The platform's revenue is a fixed percentage of the value of the restaurant's orders. The platform incurs the delivery cost. Each restaurant allocates a fixed capacity to the platform, regarding how many meals it can prepare on a given day. We model the problem as a capacitated joint assortment optimization on a bipartite graph with supply and demand nodes. We prove the problem is APX-hard while its special cases with only one restaurant or one client are NP-hard. Under some mild conditions on restaurant capacities, we develop fixed-ratio and near optimal approximation algorithms for this problem. Moreover, for the general case, using a Linear Programming relaxation of a reformulation of the original model, and by leveraging the rather soft nature of the capacity constraints in practice, we devise an asymptotically optimal assortment sampling algorithm that we show may breach any capacity constraint only by a negligible probability, as the number of clients and the restaurant capacities grow. In collaboration with one of North America's largest office meal delivery platforms, we test the performance of our LP-based sampling algorithm in a carefully controlled field experiment. Our results suggest that our methodology improves the platform's per-employee profit and revenue by at least 14%, and 10%, respectively.

2 - Exploration Optimization for Dynamic Assortment Personalization Under Linear Preferences**Sajad Modaresi, UNC Chapel Hill, Kenan-Flagler Business School, Durham, NC, United States**, Fernando Bernstein, Denis Saure

We study efficient real-time data collection approaches for an online retailer that dynamically personalizes assortments based on customers' attributes. We study the structure of efficient exploration in this setting, propose efficient learning policies, and prove theoretical performance lower and upper bounds. We test the performance using a dataset from a large Chilean retailer.

3 - Basic Reusability and Beyond: Joint Inventory and Online Assortment Optimization with Reusable Resources**Yukai Huang, WashU, St Louis, MO, United States**, Jacob Feldman, Heng Zhang

We introduce a joint optimization framework for deciding initial inventory levels and offering personalized assortments over a finite selling horizon under a multinomial logit choice model, when resources are reusable. In the classical setup—where each unit returns after a random duration with an increasing failure rate (IFR) distribution—we develop a constant-factor approximation by formulating a fluid relaxation that replaces IFR durations with geometric analogues. A key technical contribution is a novel comparison between the CDFs of geometric and IFR random variables. We then generalize to a network reusability model in which returned resources may transform (e.g., be damaged or relocated) and propose an iterative inventory refinement procedure that adapts to online assortment feedback. We prove strong, network-dependent performance bounds for this scheme. Numerical experiments over a variety of reusability scenarios demonstrate that our algorithms achieve near-optimal revenue, underscoring both the robustness and practical efficacy of our approach.

4 - Maximum Load Assortment Optimization-- Approximation Algorithms and Adaptivity Gaps**Marouane Ibn Brahim, Cornell Tech, New York, NY, United States**, Omar El Housni, Danny Segev

Motivated by applications such as attended home delivery and preference-based group scheduling, where decision makers aim to steer many customers toward choosing the same alternative, we introduce a novel class of assortment optimization problems called *maximum load assortment optimization*. In these settings, given a set of substitutable products, we face a stream of customers, each deciding between selecting a product from an offered assortment or leaving without making a selection. Assuming these decisions follow the multinomial logit choice model, we define the random load of a product as the number of customers who select it. Our objective is to offer assortments to maximize the expected *maximum* load across all products. We consider both static and dynamic formulations. In the static setting, a single assortment is used throughout customer arrivals. In the dynamic setting, assortments are personalized based on all information available at the time. Both formulations present significant computational and analytical challenges. Our main contribution is to develop efficient algorithms to compute near-optimal static and dynamic policies. Specifically, we propose a polynomial-time approximation scheme for the static problem. We also show that a policy using weight-ordered assortments yields a 1/2-approximation in the static case, and a 1/4-approximation relative to the dynamic case, establishing a constant bound on the adaptivity gap. Finally, we design an adaptive policy whose expected maximum load is within a factor $1-\epsilon$ of optimal, with a quasi-polynomial time implementation.

5 - A/B Testing Under Product Stockouts**Liting Chen, McGill University, Montreal, QC, Canada**, Maxime Cohen, Sentao Miao

This paper addresses a critical challenge in e-commerce experimentation: the bias in A/B testing on total demand caused by product stockouts. When products have limited inventory and may become unavailable during an experiment, standard randomized controlled trial (RCT) estimators systematically overestimate treatment effects. This occurs because stockouts create a dynamic interference between treatment and control groups, violating the Stable Unit Treatment Value Assumption. We develop a fluid limit analysis to characterize this bias and propose a novel Stockout-Discounted (SD) estimator that accounts for demand transfer patterns when products become unavailable. By applying appropriate discount weights based on product substitution behaviors, our estimator reduces bias while maintaining equal or lower variance compared to standard RCT approaches. We establish theoretical guarantees for our approach under general choice models without assuming specific underlying customer choice behavior. Our findings provide practical guidance for e-commerce platforms seeking accurate treatment effect estimation in inventory-constrained environments, where traditional A/B testing methods fail to account for the dynamic nature of product availability.

ME54

Building B Level 2 B216

Sustainability and Innovative Marketplaces

Invited Session

Revenue Management

Chair: Jingwei Zhang, Cornell University, Ithaca, United States

Co-Chair: Alexandra Dong, UCLA Anderson, Los Angeles, CA, 90024, United States

1 - Designing Surprise Bags for Surplus Foods**Hansheng Jiang, University of Toronto, Rotman School of Management, Toronto, ON, Canada**, Fan Zhou, Andrea Li, Joline Uichanco

Too Good To Go (TGTG) has emerged as a global leader in reducing food waste by connecting consumers with retailers' surplus food through "surprise bags." While TGTG provides both environmental and economic benefits, participating stores face significant operational challenges in determining the optimal number of bags to offer. Specifically, stores must decide on the number of surprise bags before knowing the actual surplus for the day. Additionally, key questions such as whether to include regular sale items in the bags and how to allocate food across bags are critical for balancing customer satisfaction and profitability. We explore these challenges and model the stores' decision-making process by stochastic dynamic programming. Our analysis reveals that the commonly adopted practice of evenly distributing surplus food across all surprise bags may not maximize consumer satisfaction, particularly when customer utility functions are non-concave. We

propose an alternative---offering up to two types of bags---that achieves maximal consumer satisfaction while still remaining simple to implement. Determining the optimal number of bags and the total food volume to allocate at each period is a non-convex two-stage stochastic optimization problem. To address both the analytical and computational complexities, we propose two approximation policies that are computationally efficient with intuitive managerial insights and provable performance guarantees.

Our findings provide practical guidance for stores in balancing revenue generation, customer satisfaction, and sustainability objectives, thereby enhancing the long-term viability of TGTG's mission to reduce food waste.

2 - Coordinating Perishable Product Supply Chains

Xiaoyue Yan, University of Oregon, Eugene, OR, United States, Elena Belavina

Traditional supply chain coordination strategies, such as increasing order quantities to address double marginalization, may be ineffective for perishable products. We develop a two-echelon perishable-inventory model showing that perishability creates a non-monotonic relationship between order quantity and variability propagation, driven by product expiration. This leads to the coexistence of bullwhip and reverse bullwhip effects, even without traditional bullwhip drivers.

Using real-world data on 513 perishable products across 22 retail stores, we observe that 33.8% of products exhibit a reverse bullwhip effect, where variability attenuates upstream, consistent with our theoretical predictions. Empirical analysis further confirms that product characteristics non-linearly influence variability amplification, explaining more than half of the observed variation. Through structural estimation, we find that for most perishables, reducing order quantities improves supply chain profits by mitigating variability propagation, despite exacerbating double marginalization. This generates a win-win outcome: higher profitability and lower food waste.

Our findings challenge the conventional view that higher order quantities improve coordination. For perishables, smaller orders better align incentives, stabilize supply chains, and enhance sustainability. We further propose contract designs that reduce supplier demand variability and food waste, offering new insights for designing more sustainable and efficient perishable supply chains.

3 - From Waste to Taste: Structural Estimation and Market Design in Secondary Food Markets

Alexandra Yizhuo Dong, UCLA Anderson, Los Angeles, CA, United States, Auyon Siddiq, Jingwei Zhang

Commercial food waste remains a significant challenge, with roughly one-third of all produced food wasted. In response, digital marketplaces now connect food retailers with consumers willing to purchase surplus at discounted prices. The effectiveness of these secondary markets hinges on broad seller participation. This study examines how two features of market design — price competition and the presentation of store ratings — shape seller entry decisions on a leading food waste platform.

We develop a structural model that jointly captures consumer demand and seller entry behavior. On the demand side, we combine consumer arrivals with a multinomial logit choice model that captures spatial, temporal, and demographic factors, as well as store-level features such as pricing and platform ratings. On the supply side, we model seller entry as a non-cooperative game and propose a novel discrete optimization-based estimation approach that improves both estimation accuracy and computational efficiency relative to the conventional Simulated Method of Moments (SMM).

Our empirical analysis draws on a rich dataset of hourly transactions from approximately 3,000 retailers over a three-month period, spanning four major U.S. cities. We supplement this with demographic information and data from Google Places. The results demonstrate that both pricing policies and information design meaningfully affect seller participation and overall marketplace performance. These findings offer actionable insights for platform operators aiming to increase participation, reduce food waste, and advance more sustainable retail practices.

4 - Hyperlocal P2p Food Sharing: Estimating the Effect of the Entry of the Commercial Food Donation Channel

Ekaterina Astashkina, Ross School of Business, University of Michigan, Ann Arbor, MI, United States, Masha Shunko, Yixin Iris Wang

We are examining data from the largest global peer-to-peer food sharing platform. This platform facilitates the exchange of soon-to-expire food items among platform's users. We investigate the impact of introducing a new commercial donation channel, which allows various food establishments (e.g., supermarkets, grocery stores, bakeries) to donate surplus food directly to households. Specifically, we aim to understand how this addition influences donation patterns among users of the platform.

ME55

Building B Level 3 B314

Data, Decisions and Disparate Objectives

Invited Session

Revenue Management

Chair: Giannis Fikioris, Cornell University, Ithaca, NY, United States

Co-Chair: Siddhartha Banerjee, Cornell University, Ithaca, NY, United States

1 - Magnolia: Matching Algorithms via Gnns for Online Value-to-Go Approximation

Ellen Vitercik, Stanford University, Stanford, CA, United States, Alexandre Hayderi, Amin Saberi, Anders Wikum

Online Bayesian bipartite matching is a central problem in digital marketplaces and exchanges, including advertising, crowdsourcing, ridesharing, and organ allocation. We introduce a graph neural network (GNN) approach that emulates the problem's combinatorially-complex optimal online algorithm, which selects actions (e.g., which nodes to match) by computing each action's value-to-go (VTG) -- the expected weight of the final matching if the algorithm takes that action, then acts optimally in the future. We train a GNN to estimate VTG and show empirically that this GNN returns high-weight matchings across a variety of tasks. Moreover, we identify a common family of graph distributions in spatial crowdsourcing applications, such as rideshare, under which VTG can be efficiently approximated by aggregating information within local neighborhoods in the graphs. This structure matches the local behavior of GNNs, providing theoretical justification for our approach.

2 - Marketplace Operators Can Induce Competitive Pricing

Tiffany Ding, University of California, Berkeley, Berkeley, CA, United States, Dominique Perrault-Joncas, Orit Ronen, Michael Jordan, Dirk Bergemann, Dean Foster, Omer Gottesman

As e-commerce marketplaces continue to grow in popularity, it has become increasingly important to understand the role and impact of marketplace operators on competition and social welfare. We model a marketplace operator as an entity that has both the ability to facilitate sales by third parties as well as sell the product on its own. We formalize this market structure as a price-quantity Stackelberg duopoly in which the leader is a marketplace operator and the follower is an independent seller who shares a fraction of their revenue with the marketplace operator for the privilege of selling on the platform. The objective of the marketplace operator is to maximize a weighted sum of profit and a term capturing positive customer experience, whereas the independent seller seeks solely to maximize their own profit. We derive the subgame-perfect Nash equilibrium and find that it is often optimal for the marketplace operator to induce competition by offering the product at a low price to incentivize the independent seller to match their price.

3 - Learning to Coordinate Bidders in Non-Truthful Auctions

Tao Lin, Microsoft, Cambridge, MA, United States

In non-truthful auctions such as first-price and all-pay auctions, the independent strategic behaviors of bidders, with the corresponding equilibrium notion -- Bayes Nash equilibria -- are notoriously difficult to characterize and can cause undesirable outcomes. An alternative approach to designing better auction systems is to coordinate the bidders: let a mediator make incentive-compatible recommendations of correlated bidding strategies to the bidders, namely, implementing a Bayes correlated equilibrium (BCE). The implementation of BCE, however, requires knowledge of the distribution of bidders' private valuations, which is often unavailable. We initiate the study of the

sample complexity of learning Bayes correlated equilibria in non-truthful auctions. We prove that the BCEs in a large class of non-truthful auctions, including first-price and all-pay auctions, can be learned with a polynomial number $O(n/\epsilon^2)$ of samples from the bidders' value distributions. This moderate amount of samples means that learning to coordinate bidders is statistically efficient. Our technique is a reduction to the problem of estimating bidders' expected utility from samples, combined with an analysis of the pseudo-dimension of the class of all monotone bidding strategies of bidders.

4 - Envy-Inefficiency Scaling Laws

Siddhartha Banerjee, Cornell University, Ithaca, NY, United States

In many resource allocation settings with stochastic demands and supplies, a principal must tradeoff between allocating in a way which is globally optimal but unequal across agents, or can enforce equality across allocations resulting in unavoidable efficiency losses. The specific functional form of these envy-inefficiency tradeoffs depends in subtle ways on the structure of the underlying allocation problem. I will illustrate this with three examples -- fixed supply (i.e., knapsack) settings, fixed supply with perishable items, and variable supply (i.e., inventory) settings. In each case we obtain sharp characterizations of the envy-inefficiency scaling laws, and in the process, get better qualitative insights into these various settings, as well as understand how to design allocation policies to achieve Pareto optimal points on these tradeoff curves.

Based on joint work with Chido Onyeze, Sean Sinclair, Chamsi Hssaine and Christina Lee Yu.

5 - Clock Auctions Augmented with Unreliable Advice

Daniel Schoepflin, Rutgers University, New Brunswick, NJ, United States

We provide the first analysis of clock auctions in the learning-augmented framework. These auctions satisfy a unique list of very appealing properties making them particularly well-suited for real-world applications. Early work that evaluated their performance from a worst-case analysis perspective concluded that no deterministic clock auction with n bidders can achieve a $\log^{1-1/n}(n)$ -approximation of the optimal social welfare for a constant $t > 0$, even in simple settings. This pessimistic impossibility result heavily depends on the assumption that the designer has no information regarding bidders' values. Leveraging the learning-augmented framework, we instead consider a designer equipped with some advice regarding the optimal solution; this advice can provide useful guidance if accurate, but it may be unreliable.

Our main results are learning-augmented clock auctions that use this advice to achieve much stronger performance guarantees whenever the advice is accurate, while maintaining worst-case guarantees even if this advice is arbitrarily inaccurate. Our first clock auction achieves the best-of-both-worlds: $(1+c)$ -consistency for any desired constant c and $O(\log(n))$ -robustness; we extend this auction to achieve error tolerance. We then consider a much stronger notion of consistency, which we term "consistency-infinity", and provide an auction that achieves a near-optimal trade-off between consistency-infinity and robustness. Finally, using our results, we prove bounds on the "cost of smoothness," i.e., the robustness achievable if we require that the performance of the auction degrades smoothly as a function of error.

This work is based on a paper which originally appeared at SODA 2025 with co-authors Vasilis Gkatzelis and Xizhi Tan.

ME56

Building B Level 3 B315

Data: Markets and Collaboration

Invited Session

Auctions and Market Design

Chair: Ruta Mehta, UIUC, Urbana, IL, United States

Co-Chair: Aniket Murhekar, Northwestern University, Evanston, IL, United States

Co-Chair: Pooja Kulkarni, UIUC, Urbana, IL, 61801, United States

1 - Data Monetization Through Strategic Coordination of Privately Informed Agents

Thibaut Horel, Massachusetts Institute of Technology, Cambridge, MA, United States, Alessandro Bonatti, Munther Dahleh

We consider linear-quadratic games of incomplete information with Gaussian uncertainty, in which players' payoffs depend both on a privately observed type and an unknown but common state. A monopolist data platform observes the state, elicits the players' types, and sells information back to them via (possibly correlated) action recommendations. We fully characterize the class of all such implementable Gaussian mechanisms (where the joint distribution of actions and signals is multivariate normal) as well as the player-optimal and revenue-maximizing mechanisms within this class. For games of strategic complements (substitutes), both optimal mechanisms maximally correlate (anticorrelate) the players' actions. When uncertainty over private types is large, the recommendations are deterministic and linear functions of the state and reported types but are not fully revealing. We apply our results to algorithmic pricing recommendations used by platforms such as Amazon and those challenged in *U.S. v. RealPage*.

2 - Equilibrium of Data Markets with Externality

Safwan Hossain, Harvard University, Boston, MA, United States, Yiling Chen

We model real-world data markets, where sellers post fixed prices and buyers are free to purchase from any set of sellers, as a simultaneous game. A key component here is the negative externality buyers induce on one another due to data purchases. Starting with a simple setting where buyers know their valuations a priori, we characterize both the existence and welfare properties of the pure Nash equilibrium in the presence of such externality. While the outcomes are bleak without any intervention, mirroring the limitations of current data markets, we prove that for a standard class of externality functions, platforms intervening through a transaction cost can lead to a pure equilibrium with strong welfare guarantees. We next consider a more realistic setting where buyers learn their valuations over time through market interactions. Our intervention is feasible here as well, and we consider learning algorithms to achieve low regret concerning both individual and cumulative utility metrics. Lastly, we analyze the promises of this intervention under a much richer externality model.

3 - Data-Sharing Markets: Models, Protocols, and Algorithms

Raul Castro Fernandez, The University of Chicago, Chicago, IL, United States

Data-sharing consortia unlock the latent value of siloed data across organizations, yet they often fail to materialize due to uncertainties around benefits, cost allocations, and incentives. In this work, we study an economic and algorithmic approach to incentivize the formation of such consortia. The protocol ensures every participant benefits from joining and that the consortium is sustainable only when socially valuable. A key insight of our work is the tradeoff between precision and operational cost in revenue allocation: accurate but expensive methods, such as the Shapley value, can consume so much of the value generated that they disincentivize participation or prevent consortium formation altogether. We show a simple revenue allocation algorithm that—while computationally cheaper—preserves fairness and incentive compatibility in practice, thereby leaving more revenue for participants. This tradeoff between computational exactness and allocative surplus is critical for the practical formation of data-sharing markets. We evaluate our protocol empirically and show that cheaper algorithms lead to more consortia forming, with limited impact on perceived fairness. During the talk, I will also present updates to this line of research, including opportunities for improvement. This work contributes to the growing literature at the intersection of data management, market design, and platform engineering, and suggests new opportunities for sustainable, incentive-compatible data collaboration.

ME57

Building B Level 3 B316

Michael H. Rothkopf Junior Researcher Paper Prize

Award Session

Auctions and Market Design

Chair: Rad Niazadeh, Chicago Booth School of Business, Chicago, IL, United States

Co-Chair: Christian Kroer, Columbia University, New York, NY, United States

Co-Chair: Jiaqi Lu, the Chinese University of Hong Kong (Shenzhen), Shenzhen, N/A

ME58

Building B Level 3 B317

Scheduling Operations and Optimization

Contributed Session

Contributed

Chair: Thomas Lavastida, University of Texas at Dallas, Addison, United States

1 - An Online Algorithm for Scheduling Appointments Under Endogenous No-Show Uncertainty Using Machine Learning

Maxim Frankish, Ghent University, Ghent, Belgium, Broos Maenhout

Patients not showing up for their medical appointments is a recurring issue that can negatively impact healthcare operations. Next to patient-related factors, such as no-show history and the patient's age, also schedule-dependent factors, such as access times, the time of the appointment and patient waiting times, possibly resulting from patient overbooking, directly impact the no-show behaviour. Therefore, we suggest minimising the probability of no-show upfront as also schedule-dependent factors, in combination with patient-related factors, influence the patient no-show. In this study, we develop an online appointment scheduling algorithm that leverages machine learning to predict no-show and proactively assign time slots with a low associated no-show risk. We evaluate the algorithm using real-world data from a dental department and demonstrate its effectiveness in reducing expected no-shows while maintaining a balanced schedule. The results show that integrating no-show risk into the scheduling process can provide a realistic and effective alternative to traditional overbooking practices.

2 - Data-Driven Procurement Design and Optimization: Scheduling Texas Road Construction Projects' Auctions

Hayri Alper Arslan, University of Texas at San Antonio, San Antonio, TX, United States, Alp Arslan, Dakshina De Silva

Public and private organizations frequently use auctions to procure goods and services, awarding multiple contracts at regular intervals known as letting periods. In these environments, bidders often pursue and win multiple contracts, generating cost interdependencies arising from shared resources, economies of scale, project proximity, and timeline overlaps. These interdependencies can significantly influence bidding strategies and, consequently, procurement outcomes.

This study investigates how optimizing auction schedules can leverage these cost interdependencies to improve efficiency and reduce procurement costs. Focusing on Texas road construction contracts—an ideal empirical setting due to its scale and detailed auction data—we construct project-level measures capturing material overlap, schedule overlap, and spatial proximity. Regression analyses reveal that while material and schedule overlaps generally lead to higher bids, spatial proximity encourages more competitive bidding.

To better understand strategic bidder responses, we develop a structural model incorporating contract-specific and combination-specific characteristics into bidders' decision-making processes. Using this model, we estimate equilibrium bid distributions and simulate how alternative auction schedules affect procurement outcomes.

Our simulations demonstrate that modest adjustments to auction timing—shifting individual projects one period forward or backward—reduce total expected bids by 0.7%. More substantial re-timing, allowing the 50 highest-value projects to shift across a full year, leads to a 1.4% reduction in total expected bids, translating to nearly \$40 million in savings. These findings underscore the potential for strategically designed auction schedules to significantly enhance procurement efficiency by aligning project timing and geography with bidder cost structures.

3 - Algorithms for Shipping Container Delivery Scheduling

Dimitrios Letsios, King's College London, London, United Kingdom, Anna Collins, Gueorgui Mihaylov

Motivated by distribution problems arising in the supply chain of Haleon, we investigate a shipping container delivery scheduling problem. The problem models a supplier dispatching ordered products with shipping containers from manufacturing sites to distribution centers, where orders are collected by the buyers at agreed due times. The supplier may expedite or delay item deliveries to reduce transshipment costs at the price of increasing inventory costs, as measured by the number of containers and distribution center storage/backlog costs, respectively. The goal is to compute a delivery schedule attaining good trade-offs between the two. This container delivery scheduling problem is a temporal variant of classic bin packing problems, where the item sizes are not fixed, but depend on the item due times and delivery times. An approach for solving the problem should specify a batching policy for container consolidation and a scheduling policy for deciding when each container should be delivered. Based on the available item due times, we develop algorithms with sequential and nested batching policies as well as on-time and delay-tolerant scheduling policies. We elaborate on the problem's hardness and substantiate the proposed algorithms with positive and negative approximation bounds, including the derivation of an algorithm achieving an asymptotically tight 2-approximation ratio.

4 - Heuristic Scheduling Approach for Pharmaceutical Resource Planning and End-to-End Sample Testing

Apoorva Nisal, Gilead Sciences, Oceanside, CA, United States

Pharmaceutical drug development entails extensive testing for product stability and rigorous high sample load throughput testing. Lab managers face the challenge of coordinating sample submission, scheduling studies, analysts and instruments while ensuring that the studies are performed on time. The time required for each study depends on the assay to be performed and sample specific processing. A manual sample submission and testing coordination process is replaced with an end-to-end workflow to forecast incoming studies.

The platform uses instrument data, analyst training information, other data sources such as mapping of assays to instruments, time required, etc. A heuristics-based approach is developed to generate an optimal schedule aligning trained analysts, studies, and instruments daily.

Heuristic models are rule-based algorithms that follow a set of priorities to make decisions. Scheduling and capacity modeling problems are typically modeled mathematically using the mixed integer linear programming approach. However, MILP (Mixed Integer Linear Programming) problems involve large constraints for a problem of the current size. Thus, there is a tradeoff between solution optimality, computational time and flexibility in the case of heuristics vs MILP. Both open-source solver-based MILP and a heuristic model were compared.

The heuristic model was observed to be faster compared to MILP based model and is ultimately selected to derive capacity-based insights. The integrated scheduler platform ensures that the users have the flexibility to generate on-demand schedule based on instrument and analyst availability and the prioritization of any studies as required.

5 - Single-Machine Scheduling with Cooperative Agents and Nondisjoint Job Sets

Johnson Phosavanh, The University of Sydney, Sydney, Australia, Daniel Oron

We consider a new class of multi-agent scheduling problems where some jobs are allowed to be completed by more than one agent, each with a different processing time, allowing us to represent workers with different skill sets. This problem can be considered as a multi-stage optimization problem, where the assignment of jobs to agents needs to be decided, followed by the determination of an optimal schedule or vice versa. We show that this class of problems is not strictly easier or harder than the conventional competing or nondisjoint job set scenarios, and we provide an extensive analysis of problems consisting of two agents on a single machine with various scheduling criteria.

6 - Scheduling a Single Parallel-Batching Machine with Run-to-Run Constraints

Xiajie Yi, IÉSEG School of Management, Lille, France, Fan Yang, Dries Goossens, Roel Leus

The rise of artificial intelligence has significantly increased the demand for semiconductors, intensifying scheduling challenges across complex, multi-stage manufacturing processes. To enhance yield and process stability, semiconductor manufacturers widely implement Advanced Process Control (APC) techniques, particularly run-to-run (R2R) control, which adjusts process parameters based on prior outcomes to reduce drift and variability. R2R constraints often involve time- or count-based rules, triggering qualification runs to maintain process accuracy.

This paper studies a scheduling problem involving a single parallel-batching machine with R2R constraints, motivated by practical semiconductor applications. The model accounts for incompatible job families and non-identical job sizes, where jobs from the same family can be batched within machine capacity. The objective is to minimize the total weighted completion time, which is a key performance metric linked to product holding costs. To solve this problem, we propose a time-indexed formulation and a set-covering formulation. For the latter one, we develop a column generation-based algorithm, in which the pricing problem is solved via zero-suppressed binary decision diagrams.

7 - Stochastic Parallel Scheduling with Bandit Feedback

Thomas Lavastida, University of Texas at Dallas, Richardson, TX, United States, Gerdus Benade, Rathish Das

We study the fundamental problem of scheduling jobs with stochastic sizes to minimize the total completion time for both the single and parallel machine settings. Traditionally, the size distributions are assumed to be known to the scheduler; the challenge is accounting for the stochasticity. Departing from prior work on stochastic scheduling, we explore the problem of efficiently learning to schedule when the size distributions are unknown and we are unable to directly obtain samples from it. We adopt an online bandit learning framework in which an algorithm interacts with a scheduling environment over a series of discrete time periods. In each period, the algorithm commits to a schedule for the jobs on a collection of parallel machines and then observes only the random cost of the schedule, and not individual job sizes. This so-called bandit-feedback provides limited information about the underlying (unknown) job size distributions. The objective is to minimize the total regret - the difference in expected total completion time of schedules chosen by the algorithm from that of the optimal schedule under complete information.

Standard approaches based on multi-armed bandits or linear bandits are computationally inefficient. Thus, we utilize the structure of the scheduling problem to give simple, combinatorial, and computationally efficient algorithms for both settings with nearly optimal regret. In particular, both algorithms require only constant amortized time complexity per period asymptotically, making them highly efficient. Finally, we complement our upper bounds with an almost matching lower bound for the single machine setting.

ME59

Building B Level 3 B318

Healthcare Operations: Hospitals and Organ Transplant Systems

Contributed Session

Contributed

Chair: Zihan Chai, University of Texas at Austin, Austin, TX, 78705, United States

1 - Gender Combination and Team Performance: Evidence from Heart Procedures

Dapeng Chen, Lehigh University, Bethlehem, PA, United States, Shin-Yi Chou, Xiaosong Peng

Despite growing interest in how gender composition shapes team performance, little is known about its impact in high-stakes healthcare settings. In this paper, I use inpatient discharge records for two common heart procedures - percutaneous coronary intervention (PCI) and coronary artery bypass grafting (CABG) - to investigate how the gender combinations between the attending physicians (who manage inpatient care during hospitalization) and proceduralists (who perform the procedures) affect patient health outcomes. Exploiting the quasi-random pairings of physicians and proceduralists, I find that mixed-gender teams reduce in-hospital mortality by approximately 12 percent. Further decomposition analysis suggests that female physicians outperforming their male counterparts when paired with male proceduralists, whereas no significant difference is observed when partner with female proceduralists. Additional tests show that the observed benefits are not driven by increased utilization of medical resources. Instead, the results appear consistent with communication or social norms mechanisms: the mortality gap narrows with higher patient severity and greater familiarity between physicians and proceduralists, as measured by prior collaboration and the share of male proceduralists among a physician's colleagues. However, physicians' own experience and patient demographics do not explain the observed gains. Moreover, pairing female physicians with male proceduralists lowers the likelihood of experiencing repeat surgery, which may explain part of the mortality reduction. These findings emphasize the importance of gender composition in healthcare teams and suggest potential improvement in patient outcomes with careful team assignment policies.

2 - From Reactive to Proactive: Transforming Emergency Department Full Capacity Protocols with AI and Simulation

Khalid Aram, Emporia State University, Emporia, KS, United States, Orhun Vural, Bunyamin Ozaydin, James Booth, Abdulaziz Ahmed

Emergency Department (ED) overcrowding negatively impacts the efficiency and effectiveness of healthcare systems. This study explores the transformation of ED capacity management from a reactive to a proactive approach. It integrates Discrete-Event Simulation (DES) with deep learning to test the effectiveness of proactive capacity decisions. Utilizing extensive historical ED data, the DES model evaluates a proactive Full Capacity Protocol (FCP) against the traditional reactive FCP. Preliminary results demonstrate that the proactive FCP, informed by AI-based predictions, leads to improvements in key performance indicators such as reduced ED Length of Stay, lower hourly wait counts, and decreased hold times, highlighting its potential to enhance operational efficiency and mitigate the adverse effects of ED overcrowding.

3 - A Joint Modeling Framework for Time-to-Fatigue Prediction with A Single Wearable Sensor Biomarker

Lin Lu, Fairfield University, Fairfield, CT, United States

Physical fatigue is a prevalent issue in the U.S. workforce and has resulted in reduced productivity and a higher incidence of workplace accidents. Being able to estimate time to fatigue facilitates proactive interventions before fatigue onset. Existing research has primarily focused on the state, intensity, and wearable sensor characteristics of fatigue while there remains a gap in predicting time to fatigue and identifying indicative biomarkers that can dynamically signal impending fatigue risks. In this study, we utilize simulated manufacturing work tasks and incorporate a single sensor as time-varying covariate, capturing both task and individual characteristics, into a joint modeling framework to predict individual workers' time to fatigue and evaluate its effectiveness. This study demonstrates the feasibility of joint modeling to link a single longitudinal sensor trajectory with time to fatigue, offering insights into fatigue progression and practical implications for timely interventions. It highlights the potential of online monitoring and interpretable statistical methods in enhancing occupational safety and productivity.

4 - Towards Effective Emergency Department Boarding Forecasting: a Real-World Study

Lily Poursoltan, University of California, San Diego (UCSD), La Jolla, CA, United States, Karandeep Singh, Kevin Zhu

Emergency department (ED) overcrowding and boarding pose critical challenges to patient safety, contributing to increased mortality, medical errors, and compromised care quality. While accurate short-term forecasting of ED boarding volumes can enable proactive planning and resource allocation, existing methods often underperform in practice. We address this gap through a comprehensive real-world study using two years of operational data (October 2022–October 2024) from two major UCSD Health emergency departments: La Jolla and Hillcrest. We develop a novel two-stage hybrid forecasting method that combines Vector Autoregression (VAR) to capture multivariate temporal interdependencies among covariates identified through domain experts and rigorous feature importance testing, followed by Extreme Gradient Boosting (XGBoost) to model complex nonlinear relationships for refined boarding volume predictions up to four days ahead. We implement and validate this approach through real-world deployment at both hospital sites. Our extensive field testing within UCSD Health's Mission Control command center demonstrates that the proposed method achieves the lowest RMSE and MAE—particularly in dynamic settings like the La Jolla ED—on average across forecasting horizons of up to four days, consistently outperforming classical econometric models, machine learning methods, and recent time series foundation models. The approach delivers reliable uncertainty quantification, with actual ED boarding volumes rarely falling outside the forecasted confidence intervals, thereby enabling robust surge planning. This real-world implementation has empowered Mission Control leadership to anticipate demand four days in advance, optimize staffing and bed management, and coordinate inter-facility transfers—ultimately maintaining consistent lengths of hospital stays despite growing patient volumes.

5 - Gender Differences in Workload Management: Evidence from Outpatient Services

Jiayu Chen, University of Science and Technology Beijing, Beijing, China, People's Republic of, Wei Gu, Shichen Zhang, Meng Li

Gender differences in workload management have long attracted scholarly attention. In particular, the gendered response to non-promotable workplace demands, particularly extra workload (EW)—defined as additional patient appointments beyond regularly scheduled hours—remains unclear. Female and male physicians may respond differently to EW, resulting in distinct outcomes in the efficiency-quality trade-off. In hospital contexts, EW expands patient access but may exacerbate tensions between service efficiency and care quality. This study investigates how physician gender moderates the relationship between EW and the efficiency-quality trade-off in outpatient departments. Drawing on a rich observational data set, we employ an instrumental variable two-stage least squares (IV-2SLS) strategy to address endogeneity and investigate these dynamics in depth. The results reveal that female physicians are more sensitive to EW pressures, exhibiting a stronger shift from efficiency to quality compared to their male counterparts. Specifically, female physicians allocate more time to patient interactions and diagnostic tests under high EW conditions, leading to a greater reduction in service efficiency. By contrast, male physicians tend to preserve efficiency under comparable workload conditions, suggesting a relatively lower prioritization of quality-enhancing behaviors.

These results highlight the importance of gender-sensitive workload management strategies in healthcare. By recognizing the gendered nature of workload responses, healthcare managers can more effectively address the challenges of balancing efficiency and quality in physician workloads.

6 - Long-Term Graft Survival Under the Kidney Allocation System: Paradoxical Effects of Low Estimated Post-Transplant Survival (Epts) Scores

Nani ZHOU, Shanghai Jiao Tong University, Shanghai, China, People's Republic of

Background: The U.S. Kidney Allocation System employs a "longevity matching" strategy, prioritizing high-quality donor kidneys (KDPI $\leq 20\%$) for low Estimated Post-Transplant Survival recipients to maximize patient survival. However, EPTS focuses solely on patient longevity, neglecting graft survival metrics, which may lead to systemic inefficiencies.

Methods: Utilizing data from the Scientific Registry of Transplant Recipients (2015–2024, $n = 133,213$), we employed competing risks regression and stratified analyses to evaluate the impact of EPTS and Kidney Donor Profile Index (KDPI) on long-term graft outcomes.

Results: Low-EPTS recipients ($\leq 20\%$) receiving high-KDPI kidneys exhibited significantly higher cumulative graft failure rates than high-EPTS cohorts (19.2% vs. 7.5%, $p < 0.001$), with risk inversion occurring post-800 days (adjusted HR = 0.986 per year of age, $p < 0.005$). Older recipients (> 60 years) demonstrated superior graft survival due to immunosenescence (reduced alloimmune injury), yet were systematically deprioritized under current protocols. Stratified analyses revealed consistent paradoxes across KDPI strata: low-EPTS recipients of high-quality kidneys faced the highest long-term graft attrition (adjusted sHR = 0.547–0.682, $p < 0.001$).

Conclusion: The EPTS framework's overreliance on age-centric survival prediction overlooks age's dual role in modulating immune tolerance, resulting in suboptimal allocation of high-quality kidneys. We propose: (1) integrating graft-specific predictors (e.g., HLA mismatch, donor-recipient age interaction) into EPTS; (2) adopting dynamic allocation models prioritizing middle-aged recipients (35–50 years) for optimal kidneys; and (3) establishing region-specific ethical frameworks to balance utility and equity. These reforms aim to harmonize patient survival, graft longevity, and organ stewardship amid rising global ESRD burdens.

7 - The Impact of Regulating Private Sector Profits on Dialysis Access to Care

Hedayat Alibeiki, California State University San Marcos, San Marcos, CA, United States, Michael Klein

In this work, we present an optimization model for analyzing the impact of regulating private sector profits on access to care for dialysis. We consider a dialysis system with a regulator and private sector dialysis providers. The proposed model identifies the optimal location of dialysis facilities and capacity requirements for dialysis stations at the facilities, considering patient preferences for home or facility-based dialysis, the regulator's budget and dialysis providers' profitability constraints. With travel time being an important access to care consideration, this large-scale case study resulted in over 48.5 million travel time calculations as input for the optimization model. To solve large-scale instances of the problem such as the state of California, we developed an efficient meta-heuristic algorithm.

8 - An Evaluation of Out-of-Sequence Placements in the U.S. Organ Transplantation System

Zihan Chai, University of Texas at Austin, Austin, TX, United States, Diwakar Gupta

Organ procurement organizations (OPOs) play an essential role in the U.S. deceased-donor organ transplantation system by coordinating the recovery of deceased-donor organs and implementing the Organ-Procurement-and-Transplantation-Network's (OPTN's) distribution and allocation policy. Since 2021, a series of policy changes has been associated with a sharp increase in OPOs not following the OPTN policy—specifically, making out-of-sequence (OOS) placements. Unlike cases involving patient refusal or medical ineligibility, these deviations occur at the OPO level rather than at the transplant center level and are not the result of clinical judgment. OOS organs and recipients have different characteristics relative to in-sequence (IS) organs and recipients, causing a debate in the research community about the merits of this practice. This paper's aim is to assess whether the OOS practice is detrimental or beneficial from a societal perspective, and to quantify the magnitude of its impact on measures of equity and post-transplant outcomes. Using OPTN data, we present a preliminary study in which we document distributional differences between OOS and IS organs and recipients, and estimate the effect of OOS transplants using an econometric model.

ME61

Building B Level 4 B401

Distributed Analytics for Cyber Physical Systems (CPS)

Invited Session

Quality, Statistics and Reliability

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

Co-Chair: Xubo Yue, Northeastern University, Jamaica Plain, MA, United States

1 - Causality-Informed Anomaly Detection in Partially Observable Sensor Networks: Moving Beyond Correlations

Xiaofeng Xiao, Northeastern University, Boston, MA, United States, Bo Shen, Xubo Yue

Nowadays, as AI-driven manufacturing becomes increasingly popular, the volume of data streams requiring real-time monitoring continues to grow. However, due to limited resources, it is impractical to place sensors on every variable to detect unexpected shifts. Therefore, it is necessary to develop an optimal sensor-placement strategy that enables partial observability of the system while detecting anomalies as quickly as possible. Numerous approaches have been proposed to address this challenge; however, most existing methods consider only variable correlations and neglect a crucial factor: Causality. Moreover, although a few techniques incorporate causal analysis, they rely on interventions—artificially creating anomalies—to identify causal effects, which leads to limited scalability and high computational cost. In this paper, we introduce a causality-informed Q-network approach for partially observable sensor placement in anomaly detection. By integrating causal information at each stage of Q-network training, our method achieves faster convergence and tighter error bounds. Furthermore, the trained causal-informed Q-network agent significantly reduces the detection time for anomalies, demonstrating its effectiveness for sensor placement in large-scale, real-world data streams.

2 - Zero-Knowledge Cyberattack Detection for Regulatory Compliance in Critical Infrastructure Systems

Paritosh Ramanan, Oklahoma State University, Stillwater, OK, United States, H M Mohaimanul Islam

Industrial control systems (ICS) form the operational backbone of critical infrastructure networks (CIN) such as power grids, water supply systems, and gas pipelines. As cyber threats to these systems escalate, regulatory agencies are imposing stricter compliance requirements to ensure system-wide security and reliability. A central challenge, however, is enabling regulators to verify the effectiveness of detection mechanisms without requiring utilities to disclose sensitive operational data. In this talk, we introduce a cyberattack detection framework that leverages zero-knowledge proofs to reconcile these requirements and enable provable detection guarantees while preserving data confidentiality. Our approach builds on established residual-based statistical hypothesis testing methods applied to state-space detection models. Specifically, we design a two-pronged zero-knowledge architecture that enforces (i) temporal coherence of the state-space dynamics and (ii) statistical consistency of the detection tests—allowing regulators to temporally verify alarm correctness without visibility into utility-level data. We formally analyze the soundness and zero-knowledge properties of our framework and validate its practical feasibility through computational experiments on real-world ICS datasets. As a result, our work demonstrates a scalable, privacy-preserving alternative for regulatory compliance for ICS-driven critical infrastructure networks.

3 - Sequential Decision Making for Ride-Hailing Fleet Control: A Unifying Perspective

Murwan Siddig, University of Florida, Gainesville, FL, United States

Optimal control of ride-hailing systems is a challenging problem that has garnered significant attention in recent literature. Research on this topic is diverse but remains uncoordinated and lacks a unified framework. Specifically, the choice of features, decisions, constraints, and performance metrics is inconsistent across studies, and benchmark datasets are rarely published. This lack of standardization makes it difficult to analyze the impact of different problem features and to compare the performance of competing solution methods. To address this gap, we propose a unified modeling framework for controlling a fleet of ride-hailing vehicles under stochastic demand. Our framework consolidates many problem characteristics that were previously studied in isolation, including electric vehicles and ride-pooling. Furthermore, we develop reusable benchmark instances that capture a range of spatial structures and demand distributions and make them available to the scientific community. We also present preliminary results on these benchmarks using two policies from the literature, which we adapt to our unified framework.

4 - Distributed Primal&Ndash;Dual Optimization for General Empirical Risk Minimization

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

We study primal-dual algorithms for general empirical risk minimization problems in distributed settings, focusing on two prominent classes of algorithms. The first class is the communication-efficient distributed dual coordinate ascent (CoCoA), derived from the coordinate ascent method for solving the dual problem. The second class is the alternating direction method of multipliers (ADMM), including consensus ADMM, proximal ADMM, and linearized ADMM. We demonstrate that both classes of algorithms can be transformed into a unified update form that involves only primal and dual variables. This discovery reveals key connections between the two classes of algorithms: CoCoA can be interpreted as a special case of proximal ADMM for solving the dual problem, while consensus ADMM is equivalent to a proximal ADMM algorithm. This discovery provides insight into how we can easily enable the ADMM variants to outperform the CoCoA variants by adjusting the augmented Lagrangian parameter. We further explore linearized versions of ADMM and analyze the effects of tuning parameters on these ADMM variants in the distributed setting. Our theoretical findings are supported by extensive simulation studies and real-world data analysis.

ME62

Building B Level 4 B402

Geometric Data Analysis for Complex Physical Systems

Invited Session

Quality, Statistics and Reliability

Chair: Yinan Wang, Rensselaer Polytechnic Institute, 110 8th St, Troy, 12180-3522, United States

Co-Chair: Chen Kan, The University of Texas at Arlington, Arlington, TX, United States

1 - Statistical Modeling of Surface Roughness and Its Impact on Fatigue Life in Additively Manufactured Parts

Shehzaib Irfan, Auburn University, Auburn, AL, United States, Jia Liu

Additive manufacturing (AM) has the potential to manufacture intricate parts at record speeds with fewer resources, which was not possible before using conventional manufacturing techniques. The static mechanical properties of AM parts are at par with conventional ones, but their fatigue lives are significantly shorter due to multiple factors. The most detrimental is the surface roughness, and despite advances in AM, there is no way of getting rid of it yet due to the complex dynamics of the process. This surface roughness adversely affects the fatigue life and is also a source of significant variation in it. This variation is typically caused by the stochastic nature of the surface, which can significantly vary even within parts printed using the same process conditions. The problem is further compounded by the scarcity of fatigue testing data. This work aims to address this challenge by leveraging statistical techniques that not only give accurate predictions but are also interpretable. The modeling framework will be developed by leveraging the mixed models that can take into account the global trends and the part-specific variation. This framework will be able to predict the fatigue life as well as the bounds for a given confidence level. This framework can potentially solve the data scarcity problem in studying the fatigue behavior of additively manufactured parts, hence encouraging their use in mission-critical components.

2 - Kolmogorov&Ndash;Arnold Networks-Enhanced Convolutional Autoencoders for Advanced Medical Image Analysis

Boris Oskolkov, Oklahoma State University, Stillwater, OK, United States, Chenang Liu, Yu Feng

Our work proposes a deep learning method incorporating convolutional autoencoders (CAE) with Kolmogorov–Arnold Networks (KAN) for analysis and learning representations in medical images. In particular, it utilizes the CAE's strong dimensionality reduction ability to extract meaningful, low-dimensional representations from complex, high-dimensional medical imaging data. The resultant features then feed into a KAN, an adaptive, flexible network design based on the Kolmogorov–Arnold representation theorem, capable of precisely learn complex, unknown relationships hidden within medical imaging data. By combining the expressiveness and interpretability of KANs with the robust feature extraction capabilities of CAEs, the proposed framework significantly enhances predictive performance and generalization in diverse medical imaging tasks, including diagnose assistance, preliminary signs of disease identification, and anomaly detection.

3 - Multiscale Spectral Analysis of Point Clouds for Layerwise Monitoring of Additive Manufacturing

Chen Kan, The University of Texas at Arlington, Arlington, TX, United States, Yujing Yang

Point cloud data have been increasingly collected for process monitoring and anomaly detection in additive manufacturing (AM). Extracting defect-pertinent information from point clouds, nevertheless, remains a critical challenge due to the high dimensionality and unstructured nature of the data. Existing approaches for point cloud

modeling and analysis mainly focus on the spatial domain information. They are limited in the ability to fully characterize defect-induced variations, especially subtle ones, and distinguish them from benign patterns. To fill this gap, we develop a new multiscale spectral analysis framework to extend point cloud-based defect detection from the spatial to the spectral domain. Leveraging graph signal processing, surface geometric patterns of point clouds are delineated across multiple scales. Further, an incremental one-class learning scheme is designed for adaptive layer-wise monitoring. The developed framework is evaluated through simulation and real-world case studies. Experimental results have demonstrated the effectiveness and computational efficiency of the developed framework in capturing both local and global defects in layerwise monitoring of AM.

4 - Recent Advances in Dynamic 4D Shape Modeling and 4D Printing

Michael Biehler, University of Wisconsin Madison, Madison, WI, United States, Nick Nachtmann

4D shape modeling and 4D printing represent two complementary frontiers in the design of dynamic, time-evolving structures. While 4D shape modeling focuses on the time-series modeling of 3D geometries that change over time, 4D printing refers to physical devices that exhibit such behaviors by embedding responsiveness into printed materials. Through stimuli such as heat, light, or moisture, 4D-printed objects can autonomously transform their shape or function post-fabrication.

This presentation highlights recent advances at the intersection of these two domains. We introduce novel methods for modeling, predicting, and optimizing the behavior of 4D structures, leveraging physics-based simulations, machine learning, and optimization techniques. Our approach enables the creation of adaptive, programmable systems with applications in soft robotics and beyond.

5 - Generative AI for Milling Chatter Prognosis with Imbalanced Data

Weihong Guo, Rutgers, The State University of New Jersey, Piscataway, NJ, United States, Yupei Shu, Yuebin Guo

Milling chatter is a lasting issue that significantly affects surface quality, tool life, and machine health. However, chatter prognosis, specifically the prediction of chatter occurrence and resulted surface roughness, is challenging using conventional mechanistic and simulation methods due to the complex process mechanics. This study develops a generative AI model to overcome the challenge of imbalanced training datasets due to data scarcity. Synthetic multimodal data—surface images, surface roughness, and acoustic time-series data, are generated to represent various chatter severity levels. By balancing experimental data with augmented data, the multimodal generative AI model effectively predicts diverse chatter scenarios and surface roughness.

ME63

Building B Level 4 B403

Spatial AI for Context-Aware Task Guidance and Human-Machine Teaming

Invited Session

Artificial Intelligence

Chair: Mohsen Moghaddam, Georgia Institute of Technology, 30318

1 - Multimodal AI for Context-Aware Human-Robot Collaboration

Akhil Ajikumar, Georgia Institute of Technology, Atlanta, GA, United States, Sakib Reza

How can we leverage egocentric and bodily interaction cues to build multimodal AI models that enable context-aware collaboration between humans and robots in industrial settings? We introduce a multi-modal AI system, integrated with extended reality (XR) and robotics, to address two fundamental problems in this area: (1) action understanding and (2) intent estimation, both in an online, real-time fashion. Our system processes data from RGB-D video, inertial measurements, and interaction cues to infer human intent and task progression. We develop deep learning models specifically trained for multi-modal step classification and fine-grained action recognition using annotated datasets from human-robot interaction scenarios. These models operate in a real-time sensing architecture built using Microsoft's Platform for Situated Intelligence (PSI), which enables precise synchronization and processing of heterogeneous data streams. Validated through human-robot collaborative assembly tasks, the AI system adapts robotic assistance based on inferred human state, aiming to reduce cognitive workload and improve task performance (e.g., speed, accuracy). This work highlights how multi-modal AI can be applied to support real-time, context-aware adaptation in complex task execution assisted by robots. By interpreting diverse human signals during task execution, the AI enables robotic platforms and XR interfaces to respond more intelligently and supportively.

2 - Evaluating Adaptive Spatial Interactions in Ar Through Brain-Derived Metrics

Ranjana Mehta, University of Wisconsin Madison, Madison, WI, United States

Augmented Reality (AR) training serves as a cost-effective and safe alternative to conventional instructional methods. Nonetheless, instructing novices in fundamental mid-air AR interactions poses challenges. In this study, (a) we created a reliable metric for assessing user performance across various AR interaction techniques and developed adaptation models to predict additional training needs; (b) we assessed these adaptation models through a neuroergonomics lens. We conducted two user studies, tracking motor performance and neural activity using functional near-infrared spectroscopy, to investigate the effect of the adaptation model on users' skill acquisition in poking and raycasting-based AR interactions. Our findings highlight a notable gap in users' perception and learning of mid-air AR/VR interactions, the design of these interactions, and the influence of visuo-motor behavior on user effectiveness. For example, adaptation trials for poking interactions resulted in significant gains, whereas raycasting performance—characterized by complex visuo-motor requirements—showed modest improvement. From a neuroergonomics perspective, effective neural strategies associated with expertise development exhibit increased connectivity among brain regions, while heightened activation of specific areas signifies greater cognitive demands. Our research presents a multifaceted performance metric (what happens at the hand) informed by neuroergonomic evaluation (what occurs in the brain), providing crucial insights into the spatial processing characteristics of AR interactions that can drive the creation of innovative paradigms for tailored training in AR. We posit that blending psychomotor and neuroergonomic strategies creates a robust research foundation for AR/VR interfaces that prioritize visual cues or guides to enhance visual attention and aid motor preparation.

3 - Virtual Reality Research and Learning Applications: a Human-Centered Approach

Karen Chen, North Carolina State University, Raleigh, NC, United States

Virtual reality affords visualization of spatial and temporal information in three-dimensional space and enables a plethora of educational and training opportunities that would be otherwise infeasible or impossible. Despite its attractiveness and powerful abilities, studies have reported the likelihood of virtual content overload and visual distractions. This highlights the necessity of human factors, which is a discipline that concerns the interaction between humans and technologies, where capabilities, needs, and limitations of users are considered and integrated into the design of technologies.

This presentation will discuss the use of human behavior data, workload responses, performance measures, subjective preferences, and utility of large language models in understanding virtual reality systems for supporting interactive learning and cognitive load. Human factors research methodologies including usability evaluation, user-centered design methodology, and empirical experiments will be covered. This discussion is an accumulation of knowledge from project SCALE-VR, which aims to support students in conceptualizing abstract STEM concepts, specifically size and scale, as learners tend to hold an inaccurate understanding of scale. In addition, a project on the integration of large language models and gestures in extended reality to support cognitive mapping for knowledge acquisition will also be shared. Additional open-ended discussions on creating adaptive human-centered systems based on human behavioral data or physiological responses will be included.

In sum, human-centered approach can help ensure that task demand does not exceed the capabilities of a user and thereby increase the utility of virtual reality technologies for embodied, gestured-based learning experiences enhanced by immersive, stereoscopic visualization.

4 - Context-Aware LLM Assistant for Augmented-Reality Real-Time Task Guidance: a Study on the Holoassist Dataset

Mahya Qorbani, Georgia Institute of Technology, Atlanta, GA, United States, Kamran Paynabar, Mohsen Moghaddam

With the rapid growth of data generated by augmented reality (AR) headsets and related equipment, there is a need to integrate diverse multimodal information into robust, real-time assistant systems. In this work, we propose a novel context-aware Large Language Model (LLM) assistant that fuses different data modalities, like hand action data, task steps, dialogues, and etc, as historical context for on-the-fly question answering. To explore how context affects model performance, we introduce an incremental prompting framework, where each model version receives progressively richer contextual inputs. We demonstrate the feasibility of our approach using the HoloAssist dataset, where workers receive real-time support while performing AR/VR-enabled tasks. Our experiments systematically evaluate how each modality contributes to the assistant's performance, revealing that a careful combination of these data types boosts accuracy and relevance in responses. An ablation study further highlights the individual impact of each input modality. These findings underscore the potential of multimodal integration in LLM-driven AR/VR assistance, paving the way for more intuitive, contextually aware solutions in industrial settings.

ME64

Building B Level 4 B404

Behavioral Operations: Human-algorithm Interaction

Invited Session

Behavioral Operations Management

Chair: Xinyu Liang, INSEAD, Singapore, Singapore

Co-Chair: Helen Wang, London Business School, London, United Kingdom

1 - Enhancing the Wisdom of Crowds when Experts use AI Advice

Angshuman Pal, Indiana University Kelley School of Business, Bloomington, IN, United States, Asa Palley, Ville Satopaa

Good decision making depends on obtaining an accurate forecast of an uncertain variable of interest. An effective method of combining algorithmic precision with human intuition is to provide AI advice to human experts. The collective private information may observed by an individual expert as a noisy signal, so tapping into the wisdom of the crowd by combining AI-assisted forecasts from multiple experts can further improve prediction accuracy. We demonstrate that an AI-assisted crowd can suffer from two forms of bias. First, individual experts aware of the noisiness of their prior belief would over-rely on AI advice to maximize their own accuracy, which leads to the borg effect influencing the aggregate forecast. Second, a crowd of experts with behavioral predispositions against the AI would fail to update their belief rationally and therefore under-utilize the AI advice on average, leading to algorithm aversion affecting the aggregate forecast. The resultant effect of the two mutually-opposing sources of bias impacts the AI-assisted crowd's prediction accuracy. Using a stylized Bayesian model of information, we derive an optimal pivoting procedure that re-calibrates the AI-assisted crowd forecast by shifting either toward or away from the crowd's initial prediction, depending upon the relative strengths of the two biases. We provide a prescriptive empirical pivoting method that uses historical data from similar crowds of experts to numerically estimate the optimal amount of pivoting to be applied. We experimentally examine the performance of our method under different conditions of the forecasting environment using data from an online study on Prolific.

2 - How Human-AI Interaction Affects Human-Human Interactions

Kyle Hyndman, UT Dallas, Richardson, TX, USA, TX, United States, Brandon Jones

We seek to understand how Human-AI interactions affect Human-Human interactions. We study this through the lens of the indefinitely repeated prisoner's dilemma in which cooperation is strictly dominated but socially optimal. We conduct experiments where, in a first stage, human decision makers may interact with either a human or an AI algorithm. In a second stage humans interact with other humans in an identical decision environment. We seek to uncover the effects of the initial experience on subsequent cooperative behavior.

3 - 1+1>2: Adaptive Learning System for Human-AI Integrated Forecasting

Yuxuan Zhu, Technical University of Munich, Munich, Germany, Kejia Hu, Stefan Minner

Human-AI cooperation has emerged as a promising approach for better forecasting across various domains. However, current paradigms often fall short in fully exploiting the complementary strengths of human expertise and machine intelligence, resulting in the forecast being either from the algorithm or human's judgment, rather than an effective integration. To address this gap, we propose an adaptive forecasting system that dynamically integrates algorithm-driven predictions with valuable human insights. Central to our approach is a self-adaptive learning algorithm that learns from past data and selectively incorporates rational human inputs while mitigating irrational biases. Evaluation on real-world sales data from an industry partner demonstrates that our integrated system effectively enhances forecast accuracy. Additionally, we introduce comprehensive evaluation metrics to assess human performance in the forecasting process and offer practical guidance for further improvements. Our solution presents an innovative pathway for businesses to elevate decision quality and operational efficiency.

4 - How Forced Intervention Facilitates Long-Term Algorithm Adoption

Chenshan Hu, University of Colorado Boulder, Boulder, CO, United States, Xinyu Cao, Jiankun Sun, Dennis Zhang

Using the field experiment in an online education company, we find that forced algorithm usage increases its post-experiment usage by 15.8 percentage points. We develop a theoretical model to derive empirical strategies for exploring mechanisms behind and find that learning is a key driver for long-term algorithm adoption.

5 - How Does Algorithmic Assistance Impact Human Performance in Collaboration? Evidence from Algorithm-Assisted Evaluation in Software Development

Jong Sig Chung, University of Wisconsin-Madison, Madison, WI, United States

Algorithmic assistance is increasingly deployed to assist human performance in knowledge-intensive work, yet its integration may inadvertently undermine the human collaborative processes essential for human performance in knowledge-intensive work. Focusing on the introduction of algorithms in human collaboration, this study argues that it will stifle collective discussion essential for developing their shared understanding. Therefore, it will lower human's performance. Using a stacked cohort generalized Difference-in-Differences design and data about software development projects on GitHub, the study finds that although the adoption of Continuous Integration (CI) bot within a project—algorithms that assist evaluation—accelerate evaluation decisions, they also narrow the scope of collective discussions among developers during evaluation. Ultimately, this leads to a decline in developer's performance in two essential tasks where shared understanding about the project is valuable: recognizing problem areas and exploring new improvements within the project. Plus, it increases total un-resolved problem areas within the project, possibly due to reduced developer's performance. Lastly, these adverse effects are magnified when projects receive contributions from diverse knowledge domains. These findings challenge prevailing perspectives about human-algorithm augmentation, highlighting the need for a balanced approach to integrating algorithmic assistance.

6 - Tailoring Explainable Artificial Intelligence to Reduce Algorithm Aversion and Improve Profitability

Stephanie Kelley, Sobey School of Business, Saint Mary's University, Halifax, NS, Canada, Anton Ovchinnikov, Adrienne Heinrich, Sook Yee Chong

We conduct a series of lab-in-the-field and lab experiments to investigate the preferences of real AI users concerning the tailoring of explainable artificial intelligence (XAI) and its impact on reducing algorithm aversion and improving profitability. Study #1 involves a choice-based conjoint (CBC) survey to study the effect of tailoring the XAI approach, AI model, number of features, aggregation level, and lending outcome on AI user preferences. Study #2 integrates these preference results into a series of lab-in-the-field and lab experiments to determine whether tailoring the XAI visualizations reduces AI user algorithm aversion. Study #3 then measures the impact of tailoring XAI visualizations on firm profitability. We find that AI users (i) are most swayed by the XAI approach, (ii) prefer certain combinations of XAI approaches and models, (iii) prefer to see more features, (iv) do not have a preference between individual and group aggregation levels, and (v) their preferences do not change across favourable or unfavourable model outcomes. We then observe that (vi) tailoring XAI visualizations to AI user preferences leads to reduced algorithm aversion; however, this gain is restricted to favourable model outcomes. Lastly, we find that (vii) tailoring XAI visualizations allows firms to capture more profit. All studies are conducted in a lending setting, using real loan data and models developed in collaboration with data scientists from a large institutional lender in Asia. The lab-in-the-field experiments involve real AI users—the firm's loan officers—and the results are supported by a student lab experiment.

ME65

Building B Level 4 B405

Multi-Agent Optimization in Federated Learning, Markets, and Mechanism Design

Invited Session

Artificial Intelligence

Chair: Zhengyuan Zhou, Stern School of Business, New York University, New York, NY, United States

Co-Chair: Mingxi Zhu, Georgia Institute of Technology, Atlanta, GA, 30033, United States

1 - Mechanism for Collaborative Federated Learning: Pitfalls of Shapley Value**Mingxi Zhu, Georgia Institute of Technology, Atlanta, GA, United States**

This work investigates the design of incentive mechanisms for collaborative learning platforms, where strategic agents contribute data and jointly train federated learning (FL) models. We introduce a unified Multi-Action Collaborative Federated Learning (MCFL) framework that captures the “three-body” problem -- the complex interplay between strategic agents' participation, platform mechanisms, and federated learning algorithms. Our key insight is that, the FL convergence rate is endogenously shaped by the equilibrium of agents' participation level induced by the platform proposed mechanism. We consider two widely adopted mechanisms in cooperative game: Shapley Value (SV) and Marginal Contribution (MC) and compare the agents' equilibrium and the subsequent algorithm convergence induced by these mechanisms. While the Shapley Value is often preferred due to its fairness guarantees and budget balance property, we identify a critical pitfall of SV -- it encourages agents to create duplicated fake identities and split data among them, causing slow convergence of the FL algorithm and inflated cost of communication. In contrast, although the MC mechanism is not budget-balanced, it is robust to data splitting. We also introduce the system efficiency to quantify this trade-off between surplus generated from learning and communication cost. Our finding challenges the success of SV in conventional cooperative games and identifies a critical pitfall in the context of collaborative FL.

2 - Privacy-Aware Federated Learning for Heterogeneously Sensitive Data: An Example of Cross-Region Recommendation Using LLMs**Shuang Gao, Arizona State University, Tempe, AZ, United States, Mingxi Zhu**

Global E-commerce platforms operate across diverse regions—such as Amazon Asia, Amazon Europe, and Amazon US. It raises particular challenges to privacy-preserving platform algorithms design. Firstly, different regions may have varying privacy constraints, depending on local regulations (e.g. GDPR in Europe, CCPA in US). Secondly, the platform deploys different types of data that may have varying privacy sensitivities, from private data such as personal view history to public data such as product description and reviews. In this study, we propose a novel federated learning framework that highlights the hybrid training with heterogeneously sensitive data for global platforms. Particularly, we focus on cross-regional recommendation systems. While using LLM to bridge the gap in understanding multi-lingual item attributes, the platform needs a privacy-preserving solution to finetune LLMs or prompt-tuning for the recommendation tasks. We show that this hybrid FL framework is more efficient than conventional FL and it can be extended to other contexts and platform algorithms.

3 - Learning to Bid in Non-Stationary Repeated First-Price Auctions**ZIHAO HU, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Xiaoyu Fan, Yuan Yao, Jiheng Zhang, Zhengyuan Zhou**

First-price auctions have recently gained significant traction in digital advertising markets, exemplified by Google's transition from second-price to first-price auctions. Unlike in second-price auctions, where bidding one's private valuation is a dominant strategy, determining an optimal bidding strategy in first-price auctions is more complex. From a learning perspective, the learner (a specific bidder) can interact with the environment (other bidders) sequentially to infer their behaviors. Existing research often assumes specific environmental conditions and benchmarks performance against the best fixed policy (static benchmark). While this approach ensures strong learning guarantees, the static benchmark can deviate significantly from the optimal strategy in environments with even mild non-stationarity. To address such scenarios, a dynamic benchmark—representing the sum of the best possible rewards at each time step—offers a more suitable objective. However, achieving no-regret learning with respect to the dynamic benchmark requires additional constraints. By inspecting reward functions in online first-price auctions, we introduce two metrics to quantify the “regularity” of the bidding sequence, which serve as measures of non-stationarity. We provide a minimax-optimal characterization of the dynamic regret when either of these metrics is sublinear in the time horizon. Our main technical tool is the Optimistic Mirror Descent (OMD) framework, a tool largely developed within the machine learning theory community, and is well-suited for achieving minimax-optimal dynamic regret rates in this context.

4 - Optimization of Network Inventory Systems: the (Potential) Perils of the Hindsight Approach**Sheng Liu, University of Toronto, Toronto, ON, Canada, Stefanus Jasin, Jinglong Zhao**

We consider an online retailer with multiple warehouses and multiple demand locations. We assume that the retailer uses the greedy fulfillment policy (i.e., ship from the closest and/or cheapest warehouse) and consider the performance of the initial inventory allocation given by the hindsight or the stochastic programming approach, under which the retailer can first observe all the realized demands before simultaneously making the fulfillment decisions to minimize the total cost. The hindsight approach has been proposed by recent works in the literature, and is practically appealing. While this approach has its merits and performs relatively well in some cases, in general, its limitations are not well understood. In this paper, we first show that the hindsight solution is equivalent to the optimal inventory allocation under the greedy fulfillment policy but on a misspecified (more optimistic) demand arrival distribution. This optimism is sometimes benign but can also severely backfire. We provide three necessary and sufficient conditions for the hindsight solution to be asymptotically optimal (as the unit lost sale cost grows large) under the greedy fulfillment policy.

5 - Modeling and Optimizing Endogenous Consumer Acquisition in Live-Streaming Commerce**Zhiqi Qin, Zhejiang University, Hangzhou, China, People's Republic of, Shouchang Chen, Hongpeng Wang**

Live-streaming commerce—an emerging and immersive business model—thrives on real-time interactions between celebrities and consumers. This paper uncovers an intriguing phenomenon: a celebrity's follower base is endogenous to promotional activities. We study the resulting joint-pricing and traffic investment problem in live-streaming commerce. Specifically, we propose two simple yet elegant models, the EMMNL and GEMMNL, to characterize the endogeneity of follower-size in response to pricing and traffic investment. We formulate the joint optimization problem with endogenous follower size as a non-convex program. By quantifying the marginal value of acquiring viewers, we derive the optimal traffic investment policy and offer actionable managerial guidance. This analysis further reduces the joint-problem to a pricing problem with fixed traffic investment, for which we develop tractable solution methods. Under the EMMNL-model, the pricing problem admits univariate reformulation, whereas under the GEMMNL-model, we design a Fully Polynomial-Time Approximation Scheme (FPTAS). A case study using real-world data demonstrates that ignoring follower endogeneity can lead to substantial prediction errors and profit losses. Our analysis yields several key insights: (i) A modest 8% reduction in procurement cost can double total profits, explaining the prevalence of low-cost, low-price products in live-streaming commerce; (ii) Although followers account for only 10.97% of total consumers, their contribution to profits rivals that of non-followers, highlighting their outsized economic importance; (iii) The short-term marginal-value of a follower exceeds that of a non-follower, yet this relationship reverses in the long-run—implying a strategic pivot: invest in follower acquisition for short-term gains, and in non-follower engagement for sustained long-term growth.

ME66

Building B Level 4 B406

Leveraging INFORMS Resources for Graduate Students

Invited Session

Education Outreach

Chair: Richard Staats, Retired at Retired, 11148 Byrd Dr, Mclean, VA, 22030, United States

1 - Panelist for EOC Graduate Committee Session

Gokce Palak, Wake Forest University, Winston-Salem, NC, United States

The panel will provide information in two broad areas, INFORMS resources that are of value to graduate students, and the activities and products of the Educational Outreach Committee (EOC) Graduate committee focused on being value-added to the graduate student community and their supporters.

The panel will start with brief insights by each of the panelists followed by a description of notable activities by the EOC Graduate Committee. (These will be supplemented by a handout for the panel attendees.) The panel will conclude with a robust question/answer period.

More information on the committee's work can be found here:

https://zhalindor.com/Grad_EOC/

2 - Panelist for EOC Graduate Committee Session

Richard Staats, Retired at Retired, Mclean, VA, United States

The panel will provide information in two broad areas, INFORMS resources that are of value to graduate students, and the activities and products of the Educational Outreach Committee (EOC) Graduate committee focused on being value-added to the graduate student community and their supporters.

The panel will start with brief insights by each of the panelists followed by a description of notable activities by the EOC Graduate Committee. (These will be supplemented by a handout for the panel attendees.) The panel will conclude with a robust question/answer period.

More information on the committee's work can be found here:

https://zhalindor.com/Grad_EOC/

ME67

Building B Level 4 B407

AI for Security and Resilience in Complex Networks

Invited Session

Data Mining

Chair: Tao Li, City University of Hong Kong, Yeung Kin Man Academic Building, Hong Kong, N/A

Co-Chair: Tao Li, NYU, New York City, NY, United States

Co-Chair: Quanyan Zhu, NYU, New York City, NY, United States

1 - Explainable Graph Learning for Infrastructure Resilience

Parvathinathan Venkitasubramaniam, Lehigh University, Bethlehem, PA, United States

Ensuring the resilience of critical infrastructure under natural or man-made disruptions requires both accurate prediction of system behavior under previously unseen conditions and human-understandable explanations of emerging failures. This presentation introduces two complementary approaches that use graph learning and explainable AI to strengthen infrastructure resilience. First, we present a domain-informed graph learning architecture designed to predict system performance under previously unseen conditions, such as extreme weather events, cyberattacks, or cascading failures. By integrating physical dynamics, topological features, and historical data into novel graph convolutional architectures, we demonstrate, theoretically and with practical examples, that the approach can generalize effectively to out-of-distribution scenarios, even when no prior data exists for specific conditions. Second, we introduce a post-hoc explainable AI approach that performs root cause analysis on anomalies detected within complex cyber-physical systems. This method uses causal graph extraction to identify the key interactions and dependencies across the system and derive insight into pathways responsible for abnormal system behaviors. Once an anomaly is flagged through monitoring data streams and predictive models—we show how the causal extraction framework can be used both to localize the cause of the anomaly, as well as separate sensor anomalies from process failures and cyber attacks. Both these approaches bring together graph-learning methods and domain knowledge expressed through topological features and dynamical equations, with the eventual goal to produce human-interpretable explanations that help engineers and operators diagnose failures, assess systemic vulnerabilities, and design targeted interventions.

2 - Steering Dialogue Dynamics for Robustness Against Multi-Turn Jailbreaking Attacks

Hanjiang Hu, Carnegie Mellon University, Pittsburgh, PA, United States, Alex Robey, Changliu Liu

Abstract Large language models (LLMs) are shown to be vulnerable to jailbreaking attacks where adversarial prompts are designed to elicit harmful responses. While existing defenses effectively mitigate single-turn attacks by detecting and filtering unsafe inputs, they fail against multi-turn jailbreaks that exploit contextual drift over multiple interactions, gradually leading LLMs away from safe behavior. To address this challenge, we propose a safety steering framework grounded in safe control theory, ensuring invariant safety in multi-turn dialogues. Our approach models the dialogue with LLMs using state-space representations and introduces a novel neural barrier function (NBF) to detect and filter harmful queries emerging from evolving contexts proactively. Our method achieves invariant safety at each turn of dialogue by learning a safety predictor that accounts for adversarial queries, preventing potential context drift toward jailbreaks. Extensive experiments under multiple LLMs show that our NBF-based safety steering outperforms safety alignment, prompt-based steering and lightweight LLM guardrails baselines, offering stronger defenses against multi-turn jailbreaks while maintaining a better trade-off among safety, helpfulness and over-refusal. Check out the website here <https://sites.google.com/view/llm-nbf/home>.

3 - Strategic Coordination of Heterogeneous Resources for Attack Detection

Pichamon Anukularkusol, Georgia Institute of Technology, Atlanta, GA, United States, Mathieu Dahan

We study a two-player network inspection game in which the defender coordinates sensors with potentially heterogeneous detection capabilities to detect attacks placed by a strategic attacker. We assume that there are multiple types of sensors, each determined by the detected area coverage when placed at each detector location. The objective of the defender (resp. attacker) is to minimize (resp. maximize) the expected number of undetected attacks. We analytically characterize Nash equilibria of this zero-sum game in scenarios where there is no overlapping detected area coverages upon placing sensors at any detector location, and both players possibly perform a randomized strategy. We then introduce a new solution approach derived from our analysis and validate its efficacy through computational experiments.

4 - Temporal Robustness in Discrete TIME Linear Dynamical Systems

Nilava Metya, Rutgers University, New Brunswick, NJ, United States, Arunesh Sinha

Discrete-time linear dynamical systems, including Markov chains, have found many applications. However, in some problems, there is uncertainty about the time horizon for which the system runs. This creates uncertainty about the cost (or reward) incurred based on the state distribution when the system stops. This decision is critical when one needs to understand the state of the model within finite wall clock time. For example, such models are used to model attack propagation in cybersecurity, where it is important to understand worst-case outcomes and costs within a fixed wall clock time in order to take remedial actions. Given past data samples of how long a system ran, we propose to theoretically analyze a distributional robust cost estimation task in a Wasserstein ambiguity set, instead of learning a probability distribution from a few samples. Towards this, we show an equivalence between a discrete-time Markov Chain on a probability simplex and a globally asymptotically stable (GAS) discrete-time

linear dynamical system, allowing us to base our study on a GAS system only. Then, we provide various polynomial-time algorithms and hardness results for different cases in our theoretical study, including a fundamental result about the Wassertein distance-based polytope.

ME68

Building B Level 4 B408

AI for Earth Sciences: Decision Support and Risk Management for Natural Hazards

Invited Session

Data Mining

Chair: Xu Si, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Sarhadi Ali, Georgia Institution of Technology, Atlanta, GA, United States

1 - Machine-Learning-Based High-Resolution Aftershock Catalog Reveals Fault Complexity of the 2023 Kahramanmaraş Earthquake Sequence

Phuc Mach, Georgia Institute of Technology, Atlanta, GA, United States, Zhigang Peng, Xu Si, Eric Sandvol, Ekrem Zor, Mehmet Ergin

The 2023 Mw 7.8 and 7.5 doublet earthquakes in Southeastern Türkiye caused widespread devastation and triggered complex aftershock sequences along the East Anatolian Fault system. To better characterize the evolving seismicity and underlying fault structures, we deployed over 200 SmartSolo nodal stations across major clusters and high-risk areas, recording continuous 500-Hz waveforms for more than 100 days starting in May 2023. With the collected data, we built a high-resolution earthquake catalog using enhanced machine-learning-based phase-picking and association techniques. To address the challenge of limited pre-built models for high-frequency data, we developed a deep learning-based phase-picking model tailored for 500-Hz signals. Starting with using high-confidence phase picks from the Türkiye Disaster and Emergency Management Authority (AFAD) catalog, we applied rigorous quality control to create a labeled dataset. This dataset are used for training a 500-Hz phase-picking model based on a revised PhaseNet architecture. With the retraining model, our result shows approximately 100 million picks corresponding to over 50,000 events during deployment time. This model achieved superior accuracy and phase detection compared to initial 100-Hz model, generating more than 100 million picks and over 50,000 seismic events. We also performed phase association, event location, and double-difference relocation to construct a high-resolution aftershock catalog, which reveals the intricate fault structures. The detected events also serve as templates for ongoing matched-filter-based detection. Our study highlights the power of AI-driven data mining in seismology and provides new insights into rupture dynamics and seismic hazard in tectonically active regions.

2 - Physics-Informed Machine Learning for Sub-Seasonal Prediction of Tropical Cyclone Genesis

Aren Russ, Georgia Institute of Technology, Atlanta, GA, United States, Ali Sarhadi

Tropical cyclones (TCs) are among the most destructive and costly natural disasters, causing billions of dollars in damages annually. Despite their global impacts, the physical mechanisms governing their genesis remain poorly understood. Recent advances in machine learning (ML) offer promising avenues for improving the prediction of TC formation and subsequent track evolution, particularly as traditional linear models fall short in capturing the nonlinear, chaotic nature of these systems. This research develops a physics-informed machine learning framework to predict TC genesis locations globally on sub-seasonal to seasonal timescales. The model integrates a time-dependent convolutional neural network (CNN) with a Long Short-Term Memory (LSTM) network to capture both spatial and temporal dependencies in environmental conditions. Predictor variables include key surface and atmospheric drivers obtained from the ERA5 reanalysis, while storm track data for training are drawn from the IBTrACS database. The framework is trained to reconstruct snapshots of potential genesis hotspots based on historical environmental states. The combined CNN–LSTM architecture demonstrates strong skill in learning spatiotemporal patterns associated with TC genesis, effectively linking physical precursors to eventual formation. Preliminary results indicate that the model can forecast TC genesis up to four weeks in advance, highlighting its potential as a tool for early warning and disaster preparedness in a warming climate.

3 - Tuning into Earth's Voice: Leveraging Foundation Models for Earthquake Detection

Xu Si, Georgia Institute of Technology, Atlanta, GA, United States, Zhigang Peng

Recent advances in deep learning are bringing new possibilities to seismology. Although current deep learning models have achieved success in tasks such as earthquake monitoring, they are often limited to specific regions or applications. Inspired by breakthroughs in natural language processing, where large pre-trained models like the GPT series can be adapted for a wide range of tasks, we explore whether similar approaches can benefit seismology. Since both earthquake signals and natural language involve analyzing sequential data over time, we investigated whether large language models could be fine-tuned to identify seismic events. In this study, using earthquake data from California and Japan, we present a method that adapts these pre-trained language models for earthquake detection. Our results show that this approach performs comparably to traditional deep learning methods and provides greater flexibility and adaptability. These findings highlight the potential of foundation models to advance the development of more general and adaptable AI tools for earthquake monitoring.

4 - Robust Path: Geometry and Algorithms

Hao Hao, Carnegie Mellon University, Pittsburgh, PA, United States, Peter Zhang

This work aims to connect a) the set of robust optimization solutions under varying uncertainty set size and shape (robust path) with b) the optimization paths of proximal methods for the deterministic problem. We reveal the two conceptually distinct paths are indeed geometrically similar. Leveraging this geometric view, we show that the entire robust path solution sets can be approximated via first-order method optimization paths. We apply our results for: 1) generalization of a classic result in finance. 2) fast computation of adversarially robust deep learning models.

ME69

Building B Level 4 B409

Advancing Entrepreneurial Operations and Innovation

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Morvarid Rahmani, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Zeya Wang, University of Dayton, Dayton, OH, United States

1 - AI use, Startup Funding and Performance: Evidence from British Startup Firms

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

We examine the investment patterns of venture capitalists related to AI in British startup firms. We try to answer key questions such as: Whether AI ventures attract more funding? Whether the novelty of AI affects the amount of VC funding? And whether the investment contributes to firm success?

2 - Strategies for Milestone-Driven Entrepreneurial Start-Ups in Multi-Control Settings

Zhengli Wang, The University of Hong Kong, Hong Kong, Hong Kong

New venture start-ups need to survive through multiple stages of milestone targets. This study investigates strategies in a milestone-oriented environment, by modeling a start-up with its state captured by a diffusion process. The entrepreneur can choose among multiple activities with different costs, drifts and variances. Depending on whether the process reaches an upper or lower boundary, the start-up succeeds or fails. The optimal policy's structure is defined by an efficient frontier that orders the activities by two intuitive measures: effective drift and cost-effectiveness. The results can potentially provide a foundational block in the study of entrepreneurial decision-making.

3 - Noisy Sequential & Parallel Testing of Designs: Theory and Experiment**Sreekumar Bhaskaran, SMU-Cox School of Business, Dallas, TX, United States, Sanjiv Erat**

Many decision tasks, from choosing the appropriate product design to selecting the go-to-market strategy, may be conceptualized as a search process where the decision maker evaluates multiple alternatives with the goal of selecting one. We experimentally study how the managers make these decisions in the face of measurement errors.

4 - Look Before You Leap? An Operational Analysis of Hybrid Entrepreneurship**Zeya Wang, University of Dayton, Dayton, OH, United States, Morvarid Rahmani, Karthik Ramachandran**

In this paper, we investigate the ongoing debate between fully committed entrepreneurship and the hybrid approach. Different from committed entrepreneurship, hybrid entrepreneurship involves individuals concurrently pursuing traditional employment and entrepreneurial activities. To compare the merits of these two approaches, we develop a model where the entrepreneur's learning is proportional to their exploration effort. We identify the conditions that favor the hybrid approach, as well as situations where a full commitment is optimal. Furthermore, we derive the optimal allocation of effort for entrepreneurs who choose the hybrid approach. We find that committed entrepreneurship being suitable for high potential ventures. Further, different ways of supporting entrepreneurs are not equivalent because of their effect on the entrepreneur's exploration strategy. Finally, we consider the effect of employment contracts (which may require minimum effort) and the duration of the runway. Our results reveal a nuanced view of how entrepreneurs should exert effort to test their ideas.

Monday, October 27, 5:45 PM - 6:45 PM

Building A Level 4 A411

OR/MS in an Ever-Evolving Environment: Perspectives from INFORMS Fellows

Panel Session

Keynote

Co-Chair: Mark Lewis, Cornell University, Ithaca, NY, United States

1 - Moderator Panelist**Mark Lewis, Cornell University, Ithaca, NY, United States**

OR/MS in an Ever-Evolving Environment: Perspectives from INFORMS Fellows

The INFORMS Fellows represent a set of community leaders from a broad swath of INFORMS members. Come hear from some of these leaders in our field about what they view as the current and future of OR/MS. We have gathered four INFORMS Fellows with broad range of experience and knowledge for a panel discussion. Topics may include:

1. AI in teaching, publishing
1. Connections between industry/academia
1. Reduced access to federal funding
1. The most interesting research directions in...optimization, stochastics?
1. What tools will be the most useful for students in supply chain, health care, finance, service systems?

2 - Panelist**John Birge, University of Chicago, Chicago, IL, United States****3 - Panelist****Julie Ivy, University of Michigan, Ann Arbor, MI, United States****4 - Panelist****David Shmoys, Cornell University, Ithaca, NY, United States****5 - Panelist****Philip Kaminsky, Amazon, El Cerrito, CA, United States**

Building B Level 2 B206

UPS George D. Smith Prize

Plenary Keynote

Keynote

Chair: Anne Robinson, Robinson Insights, Ottawa, ON, Canada

1 - UPS George D. Smith Prize Winner Reprise: National University of Singapore**Jussi Keppo, National University of Singapore, Singapore, Singapore, James Pang**

Established in 2013, the Business Analytics Centre (BAC) at the National University of Singapore (NUS) is a result of a collaborative effort between NUS and IBM. The Centre is the operational hub for the NUS Master of Science in Business Analytics (MSBA) program, fostering applied research in business analytics through strong industry partnerships. The MSBA program, shaped and executed with NUS's academic expertise, benefits from IBM's industrial insights, integrating theory with real-world application. As an analytics hub at NUS, the BAC leverages the strengths of both the NUS Business School and the School of Computing. Over the past decade, BAC has established a robust industry network with over 100 partners, including leading multinational corporations like Accenture, Alibaba, Amazon, AXA, DBS, Ernest & Young, Google, Huawei, IBM, Johnson & Johnson, Morgan Stanley, SAP, State Street, Swiss Re, and dynamic unicorn startups such as Ant Financial, Grab, Shopee, TigerGraph, and TikTok. These collaborations extend into educational initiatives and translational research, enhancing the practical learning experience for students. Remarkably, over 800 real-world industrial analytics/OR projects have been successfully completed by NUS MSBA students in partnership with these corporations in the past 10 years. Furthermore, BAC has cultivated an engaging alumni community of MSBA graduates, who continue to contribute to and enrich the business analytics landscape at NUS.

Building A Level 4 A412

Keynote - The UPS Advantage: Integrated Network Powered by OR & AI

Plenary Keynote

Keynote

1 - The UPS Advantage: Integrated Network Powered by OR and AI**Yentai Wan, UPS, Suwanee, GA, United States**

On a typical day, UPS delivers 22.4 million packages and documents to 10.1 million customers and collects packages from 1.6 million shippers. In 2024, it achieved total revenue of \$78 billion with 444,000 employees worldwide. UPS operates in more than 220 countries and territories with approximately 1,800 owned or leased facilities. To serve in a high-premium complex logistics customer base, it leverages a fleet of 291 aircraft it owns and another 243 it charters. UPS operates one of the world's largest civilian airlines, using 734 airports throughout the world. In addition, approximately 135,000 package cars, vans, tractors and motorcycles, including more than 19,000 alternative fuel and advanced technology vehicles are used to achieve marketing leading customer service. UPS has a rich history of operational excellence and is steeped in a culture of technology innovation. Any researcher will quickly notice the reference of Operations Research (OR) in UPS could dating way back to the 1950s and most recently, the 2016 Franz Edelman Award winner, ORION that optimizes final mile deliveries. In this talk, we will unveil the curtain of the first and middle-mile UPS network with its unique challenges and opportunities. More importantly, how UPS applies OR with supplemental AI innovations to dynamically scale our networks in a highly uncertain business environment

Tuesday, October 28, 8:00 AM - 9:15 AM

TA01

Building A Level 3 A301

Veydra.io/FICO

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Advancing Decision Intelligence with AI-Assisted Simulation: Case Studies from the Strategic Decision Lab

Alton Alexander, Veydra.io., Pleasant Grove, UT, United States

See how systems thinking and AI-assisted simulation can accelerate decision-making in education and practice. Through case studies in workforce scheduling, healthcare capacity, supply chains, and training pipelines, we show how the Strategic Decision Lab enables participants to design, calibrate, and compare scenarios rapidly. Attendees will see how AI assistants support model building, scenario exploration, and insight generation while maintaining transparency and rigor. Participants will gain practical methods to integrate AI-assisted simulation into their teaching or professional practice, learning how to enrich curricula, engage learners, and apply collaborative systems thinking for real-world decision support.

2 - Reducing Time-to-Value with Fico Xpress Optimization Suite and Gen AI

Carlos Zetina, FICO Xpress Optimization, Québec, QC, Canada

Time to Value measures how long it takes for technology projects to begin providing business value. Three critical milestones to achieve this are model creation, validation, and deployment. In this showcase, we'll show how FICO Xpress Optimization Suite shortens Time to Value and supercharges Return on Investment (ROI) by providing a unified platform for development, validation, and deployment. We'll show how FICO's suite of products and Gen AI come together to turbocharge building and adoption of enterprise optimization applications deployable on any tech stack, be it server-less or via virtual machines, hosted on any cloud provider or on-premises infrastructure. Hear from Rusty Burlingame from Southwest Airlines, on how Southwest is using Xpress to improve customer experience, enhance employee engagement, and streamline operations to maximize revenue and support growth.

TA02

Building A Level 3 A302

ChiAha/GAMS

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Democratizing Digital Twins: From Real-Time Operations to Strategic Planning

Andrew Siprelle, ChiAha, Maryville, TN, United States

Digital twins serve different problem domains requiring different technical approaches. Traditional 3D kinematic simulation excels at spatial problems. Strategic capital investment decisions require different capabilities: rapid scenario analysis, statistical validation, and system interdependency revelation. This session presents a case study demonstrating time-compressed discrete event simulation methodology delivering strategic intelligence through rapid construction and statistical parameterization from historian data. The case reveals a critical insight: two equipment problems with similar historical downtime losses deliver 1.3x versus 5x actual gains when fixed. This gap illustrates why traditional loss analysis misleads prioritization decisions. The methodology enables Fortune 500-standard predictive modeling to become accessible.

2 - An Introduction to Modeling with GAMSPy

Adam Christensen, GAMS Development Corp, Fairfax, VA, United States, Steven Dirkse

Our showcase offers a hands-on introduction to GAMSPy. GAMSPy combines the high-performance GAMS execution system with the flexible Python language, creating a powerful mathematical optimization package. It acts as a bridge between the expressive Python language and the robust GAMS system, allowing you to effortlessly create complex mathematical models and applications. Join us to explore GAMSPy's fundamental functionalities through practical, interactive exercises. We'll cover everything from defining sets, parameters, variables, and equations to solving models and retrieving results, all within a familiar Python environment. Beyond the basics, we'll also provide a glimpse into more advanced features, demonstrating how GAMSPy can streamline complex modeling workflows and enhance your analytical capabilities. Whether you're a seasoned GAMS user looking to integrate with Python or a Python user curious about optimization, this workshop will equip you with essential skills needed to get started and demonstrate what is possible with GAMSPy.

TA03

Building A Level 3 A303

Research in OM-Finance Interface

Invited Session

MSOM: iForm

Chair: Jing Wu, Chinese University of Hong Kong, Hong Kong, Hong Kong

Co-Chair: Yilin SHI, The Chinese University of Hong Kong, Hong Kong, N/A

1 - Deep Hedging Strategies for a Risk-Averse Hog Farm

Ye Liu, Syracuse University, Syracuse, NY, United States, Panos Kouvelis, Yunzhe Qiu, Danko Turcic

We employ machine learning techniques to develop a comprehensive, integrated risk management model tailored to the operations of a livestock farm. This model is designed to assist a risk-averse farmer in making optimal decisions under uncertainty by dynamically integrating production and financial risk considerations. Specifically, it provides guidance on both the timing and method of selling livestock, taking into account market conditions, price volatility, and the farmer's risk preferences. In addition, the model formulates effective hedging strategies using available financial instruments to mitigate exposure to adverse price movements. By leveraging data-driven methods, the framework adapts to changing environments and enhances the farmer's ability to manage risk in a systematic and forward-looking manner.

2 - Buying Under Fire: Inflation Ambiguity and Consumer Behavior in Hyperinflation Crisis

Haoran Yu, Syracuse University, Syracuse, NY, United States, Burak Kazaz, Scott Webster

In the context of hyperinflation, governments often underreport official inflation figures to conceal economic policy failures. This systematic understatement erodes public trust in official statistics and motivates independent organizations to publish alternative estimates. In Türkiye, for example, the inflation rates released by independent institutions are typically higher than those reported by the government. The coexistence of divergent inflation figures introduces significant ambiguity for consumers, complicating their perception of the economic environment. This paper investigates how both the level of official inflation rate and the ambiguity surrounding the true inflation rate affect consumer purchasing behavior. We leverage detailed sales data from Kombucha 2200, the largest kombucha producer in Türkiye, and compile monthly inflation rates reported by both the Turkish Statistical Institute (TURKSTAT) and independent sources such as Istanbul Chamber of Commerce (ITO). Our analysis reveals that higher reported inflation rates lead to increased consumption, consistent with intertemporal substitution under anticipated price hikes. Conversely, actual price increases reduce demand, as expected. Most notably, we provide the first empirical evidence that inflation ambiguity—captured by the discrepancy between government-declared and independently estimated inflation rates—negatively affects consumer demand. This finding suggests that when consumers face uncertainty about the true inflation rate, they may delay or reduce purchases due to heightened economic anxiety or distrust in market signals. Our results highlight the importance of credible and transparent inflation reporting, particularly in hyperinflationary economies, where informational integrity plays a critical role in sustaining consumer confidence and market stability.

3 - Contract Mechanisms to Enhance Compliance in Coffee Futures Under Price and Yield Uncertainty

Mariana Losada-Agudelo, Rensselaer Polytechnic Institute, Troy, NY, United States, Sebastian Souyris, M. Hakan Hekimoglu

Agriculture faces significant uncertainties due to volatile markets and unpredictable climate conditions that affect crop yields. Among agricultural commodities, coffee is particularly susceptible to price volatility, with recent price spikes driven by low yields caused by adverse weather. Futures contracts have been a common strategy to mitigate price volatility, providing financial stability to coffee farmers. However, the lack of flexibility in these contracts, continued price volatility, and limited financing have created incentives for non-compliance, undermining their effectiveness. This research proposes an analytical model that uses incentives and penalties to improve futures contract compliance by enabling farmers to make decisions under price and yield uncertainty. We design mechanisms that include upfront payments and penalties, aiming to balance risk between parties and reduce the likelihood of strategic default. The upfront payments provide farmers with early liquidity, allowing them to invest in their crops and improve productivity by addressing financial constraints. In turn, the penalty component discourages deviation after contract acceptance by increasing the cost of non-compliance once commitments have been made. The contracts aim to ensure product delivery while optimizing the financial stability of farmers. This study contributes to the literature on agricultural contract design, demonstrating how appropriate incentives can reduce the risk of supply chain disruptions and improve cooperation between farmers and buyers.

4 - Supply Chain Washing: Strategic Disclosure of Corporate Suppliers

Yilin SHI, Chinese University of Hong Kong, Hong Kong, China, People's Republic of, Jing Wu, Yu Zhang, Yuqing Zhou

This paper examines how firms manage their disclosure of customer-supplier relationships to create a favorable green image in their supply chain. Our study finds robust evidence that firms strategically disclose relationships with environmentally responsible ("good") suppliers while withholding relationships with "bad" suppliers, *ceteris paribus*. Our analysis suggests that this selective disclosure is driven by opportunistic supply chain greenwashing behavior rather than by a genuine effort to communicate improvements in ESG performance through partnerships with high-ESG suppliers. Further, we document that firms are more likely to feature green suppliers in the 10-K's Business section and ESG reports, while less environmentally responsible suppliers tend to appear in the 10-K's Risk Factors section. Moreover, the extent of this strategic disclosure decreases as regulations on environmental information transparency become more stringent. Additionally, we observe that companies that strategically disclose "green" suppliers have higher stock returns and asset turnover in the short term, although these effects diminish over time. This suggests that it takes time for investors and consumers to fully understand such strategic disclosures.

5 - Credit Chain and Sectoral Co-Movement: a Multi-Region Investigation

Hsiao-Hui Lee, National Chengchi University, Taipei, Taiwan, S. Alex Yang, Yuxuan Zhang, Kijin Kim

This paper empirically examines the relationship between sectoral co-movement and trade credit usage in a multi-region context. Building on the models of Shea (2002) and Raddatz (2010), we develop a framework that captures the effects of trade credit on sectoral co-movement, both within countries and across borders. Using the Multi-Regional Input-Output Table developed by the Asian Development Bank, we assemble a dataset covering 14 manufacturing industries across 53 countries. Our findings provide empirical evidence that trade credit is a key channel for propagating both domestic and cross-border shocks, amplifying their impact on global industries. We show that the influence of domestic credit chains on sectoral co-movement is twice as strong as that of international chains. Furthermore, we analyze two outsourcing strategies—friend-shoring and nearshoring—under current geopolitical tensions, and find that these strategies may actually exacerbate shock propagation.

TA04

Building A Level 3 A304

Corporate Policy, Governance & Decision Making

Contributed Session

Contributed

Chair: Cathleen Johnson, West Virginia University, Reynolds Hall, Morgantown, WV, 26502, United States

1 - Stochastic Income and Optimal Policies: a New Analysis

Seyoung Park, University of Nottingham, Nottingham, United Kingdom, Alain Bensoussan, Adannah Duruoha, Viswanath Ramakrishna

This paper provides a new analysis for optimal consumption and investment policies with stochastic income. Our analysis gives some kind of procedure for solving the Bellman equation with the dimension reduction scheme developed. In particular, we provide an interpretation of the reduced problem based on the probabilistic approach. The value function and optimal strategies are all explicitly characterized with analytic comparative statics provided, which is helpful to understand much of the economics in the paper. We also develop with the convergence theorem a numerical algorithm for policy iteration and solve the Bellman equation as a sequence of solutions to ordinary differential equations. The decision to invest more or fewer equity in the stock market in the presence of stochastic income is influenced, to a large extent, by the way in which how high risk aversion and income-to-wealth ratio are exhibited. Finally, we provide more economic forecasts with borrowing constraints and various stochastic income scenarios.

2 - Liability-Driven Investment Under Inflation Risk and Regime Switching

MYUNGJUN KIM, Pohang University of Science and Technology, Pohang-si, Korea, Republic of, Bong-Gyu Jang, Hyeontae Jo

This paper investigates the optimal liability-driven investment (LDI) strategy for pension funds to manage explicit inflation risk while the financial market switches across economic regimes. Inflation rates, real interest rates, risk premia, and volatility are modeled as regime-dependent and evolve according to a Markov process. We cast the manager's problem in continuous time: maximize the expected utility of the funding ratio (asset-to-liability ratio), subject to a Value-at-Risk (VaR) constraint. By combining an option-based hedging framework with the regime-switching dynamics, we find the optimal portfolio weights for the inflation-linked bond, the risk-free asset, and the stock in every regime.

Numerical results highlight three distinctive patterns. First, in high-inflation regimes, the asset allocation tilts sharply toward the inflation-linked bond. Second, in low-inflation and low-volatility regimes, a larger equity investment is optimal if the plan is not severely under-funded. Third, in high-volatility regimes the VaR bound activates quickly, triggering rapid de-risking. These findings demonstrate that a regime-aware, inflation-conscious LDI policy can systematically adapt to macroeconomic conditions and improve the stability of pension funding ratios.

3 - Measuring Ceo Real Options Mindset Using Textual Analysis: Corporate Policy Implications

Oupin Tang, Xi'an Jiaotong-Liverpool University, Suzhou, China, People's Republic of, Tarik Driouchi, Lenos Trigeorgis

This paper studies the decision making and corporate policy implications of CEO real options mindset (ROM) in a large dataset of US-listed firms. We elicit CEOs' degrees of real options reasoning from corporate earnings calls through three key behavioral real options attributes (analytical thinking, forward-looking long-term orientation, and adaptability) and develop a composite score for ROM at the firm level. We find that high ROM CEOs are more risk averse but more managerially able than low ROM counterparts. This in turn influences firm performance outcomes, financial policy, innovation, investment, and governance decisions. Firms with higher ROM scores hold more growth options and option-embedded intangible assets, have lower dividend payouts, show higher organizational flexibility and investment efficiency, display higher skewness in market returns, and have more diverse boards and lower CEO relative pay.

4 - Optimal Inventory Control Under Stochastic Price with a Sell-Back Option

Peter Berling, NTNU, Trondheim, Norway

Stochastic purchase prices are a reality in many industries, e.g. metal manufacturing. It has been shown that a price dependent order-up-to policy is optimal under these circumstances. However, if the material can be sold back at market price this policy must be complemented with a price and location dependent sell-down-to policy. In this work we show the structure of such a policy and the money left on the table if the purchasing decision is made without accounting for the possibility to sell-back the material at a future date based on the observed demand and price.

5 - Optimal Information Disclosure with Rationally Inattentive Receivers

Bernardo Quiroga, West Virginia University, Morgantown, WV, United States, Mauricio Sauma-Webb

A privately informed receiver is being persuaded by a sender to take a given course of action. After the persuasion stage, the receiver may decide how much information to purchase from a private information technology. With this information at hand, the receiver must then choose between following the sender's recommended action or not. In this context, we characterize the receiver's information acquisition strategy and the sender's disclosure policy. We show that when the receiver is rationally inattentive, the sender will completely crowd out information acquisition. However, information acquisition may arise in equilibrium when the receiver's information structure is restricted.

6 - Time Preferences are Not Income Dependent: Evidence from the Lab

Cathleen Johnson, West Virginia University, Morgantown, WV, United States, Antonello Maruotti, Bernardo Quiroga

Using data from a lab-in-the-field experiment with 881 participants across 13 Canadian cities and communities, we examine the relationship between income and willingness to delay financial gratification. Participants faced 40 opportunities to choose between immediate and delayed cash payoffs.

Our findings show that, aside from those in the top income decile (earning over \$100,000 in 2022 CAD), participants—regardless of income level, including those classified as very poor (earning less than \$15,340 in 2022 CAD)—displayed similar levels of patience.

Mathematical numeracy is a strong, consistent predictor of patience. However, the interaction between income and numeracy does not enhance predictive power. Additionally, individuals with higher debt levels tend to be less patient. Only the wealthiest participants demonstrated a significantly higher propensity to wait for larger, later rewards. Income level does not predict relative patient time preferences for the majority of the population.

TA05

Building A Level 3 A305

Service Science Award Session

Award Session

Service Science

Co-Chair: Renyu Zhang, The Chinese University of Hong Kong, 12 Chak Cheung Street, Hong Kong, N/A

Co-Chair: Muge Yayla Kullu, University of Texas at San Antonio, San Antonio, TX, United States

1 - Bundling OR Discounting? Field Experiments for Healthy and Unhealthy Food

Nymisha Bandi, Stony Brook University, Stony Brook, NY, United States, Maxime Cohen, Saibal Ray

Retailers routinely face the strategic decision of how to promote their products. Two of the most common promotion strategies are bundling and discounting. Determining which strategy is more effective, and how their effectiveness varies across product categories, is an open question in the world of retail. To address this, we conducted a multi-store field experiment with a global convenience store chain that regularly uses bundle promotions (particularly for certain unhealthy snacks). Our study focused on healthy and unhealthy food categories to understand how these promotions influence consumer behavior and financial outcomes. Specifically, we tested these strategies--- discounts and bundles---applied to (i) healthy snacks, (ii) unhealthy snacks, and (iii) both categories simultaneously. For healthy snacks, we found that while discounting was more effective at boosting sales, bundling was more effective at increasing revenue and profit. Additionally, both promotion strategies for healthy snacks increased the sales of other unhealthy items, as consumers may engage in moral licensing, using their healthy purchases as justification to indulge in less healthy options. In contrast, neither promotional strategy had an effect on sales for unhealthy snacks. Based on this surprising finding, we convinced the retailer to conduct a follow-up experiment to investigate the impact of discontinuing a long-standing bundle promotion on unhealthy snacks. The results of this second experiment revealed no change in sales but an increase in revenue and profit, confirming our initial conclusion. These insights provide valuable guidance for retailers seeking to optimize promotion strategies for different product categories.

2 - The Value of Last-Mile Delivery in Online Retail

Zhikun Lu, NYU Shanghai, Shanghai, China, People's Republic of, Ruomeng Cui, Tianshu Sun, Lixia Wu

Last-mile delivery, the most expensive stage of the shipment process, has become increasingly important in online retail, raising the strategic question of whether firms should outsource it to customers via pickup stations or offer home delivery. In this paper, we examine the economic value of last-mile delivery to inform this high-stakes decision. Partnering with Cainiao, Alibaba's logistics subsidiary, we exploit a quasi-experiment where home delivery was sequentially rolled out to pickup stations in 2021. A staggered difference-in-differences design reveals that home delivery significantly increases sales and customer spending on Alibaba's retail platform. Because last-mile delivery is labor-intensive and capacity constrained, effectively allocating delivery resources is crucial. To address this challenge, we propose a novel operations-aware targeting framework that integrates causal machine learning with constrained optimization to identify and prioritize the most valuable customers. The framework incorporates both capacity and fairness constraints to maximize profits while maintaining equitable service access. We further extend it with routing optimization to enable value-aware delivery planning that accounts for both sales uplift and spatial efficiency. Our results demonstrate that targeting and routing policies informed by causal machine learning enables firms capture these gains more effectively. Taken together, the findings suggest that last-mile home delivery is not merely a cost center but also a revenue driver. Online retailers can unlock substantial value by adopting tailored logistics strategies that balance customer needs with resource constraints.

3 - A Structural Analysis of Freight Delays in the Indian Railway Network

Kashish Arora, CUHK Business School, Hong Kong, Hong Kong, Himanshu Arha, Milind Sohon, Milind Sohon, Raja Gopalakrishnan

Despite being cost-effective and sustainable, rail freight has been steadily losing market share globally, largely due to slow speeds. In Indian Railways, the world's fourth-largest network, average freight speed has stagnated at 25 kmph for decades, mainly because passenger trains are prioritized by section controllers operating under real-time constraints. This paper examines the causes of freight delays by modeling section controllers' stop-and-hold decisions using high-frequency network-level data.

We estimate a structural model of controller decision-making that accounts for spatial correlation in unobservables (via a copula-based approach) and corrects for endogeneity using a network-based instrumental variables strategy. The model quantifies key preferences—such as prioritizing passenger trains, queue spillovers, and strategic holds at large stations—and enables evaluation of policy interventions.

We simulate infrastructure-heavy and operational counterfactuals. A Freight-Only Corridor (FOC), modeled as a network without passenger traffic, reduces delays by 37% and increases speeds by 16%. Enhancing this with design features from India's planned Dedicated Freight Corridor further boosts benefits.

We also evaluate two low-cost operational policies: a 60-minute dispatch threshold reduces dwell times by 34% and increases speeds by 14%, while vertically stacking freight to consolidate 30% of capacity raises speeds by 8%. These approaches, while accounting for real-world constraints such as passenger interference and consolidation delays, deliver benefits comparable to FOCs at significantly lower cost.

Our findings offer actionable insights and a policy-relevant menu of infrastructure and operational strategies to improve freight movement.

4 - Why the Best Machine May Not be the Best: Incentivizing Human-Machine Collaboration

Anyan Qi, The University of Texas at Dallas, Richardson, TX, United States, Xiaotong Guan, Shouqiang Wang

As AI reshapes our world, robust human-machine collaboration is paramount to modernizing crucial operational decisions such as demand forecasts. In this paper, we study how the machine should be designed to incentivize human contributions. In our setting, a principal designs a forecasting algorithm by specifying its precision and delegates the final forecasting task to an agent. After receiving the machine's forecast, the agent decides whether to acquire an additional private signal before making the final forecast. We find that the principal may optimally reduce the machine's precision below its technological limit—even when enhancing such precision within that limit is costless. Essentially, the machine's precision not only serves a functional role but also acts as an incentive instrument to motivate the agent to gather additional information. Interestingly, the agent may over-acquire more information, resulting in possibly higher forecasting accuracy, than what would be optimal if the forecasting task were not delegated. On an optimistic note, we find that humans and machines are complementary: the better the information the agent can acquire, the more precise the machine should be.

Keywords: forecast, information acquisition, information design, principal-agent model

TA06

Building A Level 3 A306

Learning and Optimization through Adaptive Experimentation

Invited Session

MSOM: Service Operations

Chair: Wenjia Ba, University of British Columbia, Vancouver, BC, Canada

Co-Chair: Yueyang Liu, Rice University, Houston, United States

1 - A Behavioral Model for Multi-Armed Bandits: Thompson Sampling Meets Quantal Choice

Yifan Feng, National University of Singapore, Singapore, Singapore, Jingying Ding, Ying Rong

How do people navigate the exploration-exploitation (EE) trade-off when making repeated choices with unknown rewards? We study this question through the lens of multi-armed bandit problems and introduce a novel behavioral model, *Quantal Choice with Adaptive Reduction of Exploration* (QCARE). It generalizes Thompson Sampling, allowing for a principled way to quantify the EE trade-off and reflect human decision-making patterns. The model adaptively reduces exploration as information accumulates, with the reduction rate serving as a parameter to quantify the EE trade-off dynamics. We theoretically analyze how varying reduction rates influence decision quality, shedding light on the effects of "over-exploration" and "under-exploration." Empirically, we validate QCARE through experiments collecting behavioral data from human participants. QCARE not only captures critical behavioral patterns in the EE trade-off but also outperforms alternative models in predictive power. Our analysis reveals a behavioral tendency toward over-exploration.

2 - Sequential Experimental Design for Contextual Linear Optimization: a Prediction-Uncertainty-Based Approach

Beichen Wan, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States, Mo Liu, Paul Grigas, Zuo-Jun Shen

We study the problem of data collection for training decision-focused predictive models, allowing for a wide range of loss functions beyond the standard squared loss. In our setting, the learner can sequentially design experiments—selecting which samples to label—with the goal of building a well performing prediction model using as few labeled data points as possible. However, the complexity of general loss functions makes computing optimal query strategies challenging. To overcome this, we introduce a new algorithm that leverages local prediction uncertainty within a confidence set of candidate predictors. This approach enables efficient, tractable decision-making during the data acquisition process. We establish theoretical guarantees by deriving an upper bound on the label complexity of our method, showing its advantage over conventional supervised learning approaches in contextual stochastic linear optimization (CSLO) settings.

3 - Selective Randomization Inference for Adaptive Designs

Zijun Gao, University of Southern California, Los Angeles, CA, United States

Many clinical trials are structured with multiple stages, where data analysis is conducted after each stage to inform subsequent participant recruitment and treatment allocation. This adaptive approach allows for early elimination of ineffective treatments or targeted recruitment of subpopulations showing potential benefits. Analyzing such trials presents challenges as the data is utilized twice: first for selecting the design and null hypothesis, and then for testing the chosen hypothesis using the data generated under the selected design. Classical statistical methods are inadequate as they require pre-specified data generating mechanisms and null hypotheses. Existing solutions are often limited in scope, tailored to specific designs. In this work, we propose a general framework capable of handling diverse designs and adaptive choices. Our approach leverages post-selection inference principles to develop a selective randomization p-value. Notably, it does not necessitate assumptions about the distribution of outcomes or covariates, or the dependency structure among participants. We demonstrate that our method enhances statistical power compared to other valid tests while maintaining control over the selective type-I error in simulated data and hypothetical clinical trials.

4 - Distributional Robust Experiment Design

Guangyan Gan, Nanyang Technological University, Singapore, Singapore, Tu Ni, Zhenzhen Yan, Jinglong Zhao

Data-driven experiment design has attracted significant attention in recent years. In this paper, we introduce, for the first time, a distributionally robust framework for experimental design. We investigate the relationships between optimal decisions under the minimax, sample average approximation (SAA), and distributionally robust optimization (DRO) paradigms. Furthermore, we characterize the conditions under which the DRO framework outperforms SAA under both Bernoulli design and completely randomized design. The performance of our proposed methods is evaluated through extensive numerical experiments.

TA07

Building A Level 3 A307

Non-convex Optimization

Job Market Showcase

Job Market Showcase

Chair: Yankun Huang, Arizona State University, 101 W 5th St Apt 3043, Tempe, 85281, United States

1 - The Optimal Smoothing of Sublinear Functions and Convex Cones**Thabo Samakhoana, Johns Hopkins University, Baltimore, MD, United States, Benjamin Grimmer**

A nonsmooth objective function can be optimized at an improved rate by optimizing its smooth approximation at an accelerated rate. Given that better smoothings lead to improved convergence guarantees, it is natural to seek the best possible smoothing. In this work, we provide a theoretical study of optimal smoothings of functions and sets. In particular, we completely characterize the set of optimal smoothings of sublinear functions and cones. As an application of our sublinear smoothings, we obtain well-motivated smoothings for nonsmooth composite functions. Similarly, our conic smoothings provide well-motivated smoothings for conic programming constraints.

2 - Inexact Moreau Envelope Lagrangian Method for Non-Convex Constrained Optimization Under Local Error Bound Conditions on Constraint Functions**Yankun Huang, Arizona State University, Tempe, AZ, United States, Qihang Lin, Yangyang Xu**

In this paper, we study the inexact Moreau envelope Lagrangian (iMELA) method for solving smooth non-convex optimization problems over a simple polytope with additional convex inequality constraints. By incorporating a proximal term into the traditional Lagrangian function, the iMELA method approximately solves a convex optimization subproblem over the polyhedral set at each main iteration. Under the assumption of a local error bound condition for subsets of the feasible set defined by subsets of the constraints, we establish that the iMELA method can find an ϵ -Karush-Kuhn-Tucker point with $\tilde{O}(\epsilon^{-2})$ gradient oracle complexity.

TA08

Building A Level 3 A310

Digital Economy, AI, and Work

Invited Session

Information Systems

Co-Chair: Xue Guo, Georgia State University, Atlanta, GA, 30303, United States

1 - Stimulating Feedback Contributions Using Digital Nudges: a Field Experiment in a Real-Time Mobile Feedback Platform**Xue Guo, Georgia State University, Atlanta, GA, United States, Michael Rivera, Guohou Shan, Liangfei Qiu**

Organizations increasingly recognize the value of real-time feedback for employee performance, yet motivating employees to provide sufficient high-quality feedback remains challenging. Our study explores digital nudges as tools to enhance employee participation in feedback applications. Leveraging Cognitive Load Theory and Job Characteristics Theory, we focus on two critical dimensions: timing of nudges (morning vs. afternoon) and task significance messaging. We propose that morning nudges lower cognitive load, facilitating better processing, while task significance messaging emphasizes the meaningfulness of feedback, increasing motivation. Through a field experiment within a real-time feedback application, we assessed how these dimensions influence feedback quantity and quality individually and interactively. Our findings reveal that task significance nudges significantly increase both feedback contributions and feedback quality compared to generic prompts. Further interaction analysis shows timing amplifies task significance effects by reducing cognitive load, allowing greater resource allocation for message processing. Notably, timing alone was influential only when paired with task significance messaging. Additionally, through a randomized online experiment, we investigated underlying cognitive mechanisms. Results indicate that optimal timing reduces intrinsic and germane cognitive loads, whereas task significance primarily elevates germane load and external motivation. Thus, while task significance predominantly motivates feedback contributions, timing significantly enhances this effect. This research provides theoretical insights into the cognitive and motivational processes involved in digital nudging. Practically, it guides organizations in designing more effective digital feedback interventions by strategically combining timing and task significance messaging.

2 - Prompting Entrepreneurship Evaluation: Exploring Anchoring Bias with LLMs**Lauren Rhue, University of Maryland, College Park, MD, United States, Jessica Clark**

Using an experiment, the study explores whether large language models (LLMs), like OpenAI's GPT-4o, exhibit cognitive biases when evaluating project descriptions on the crowdfunding platform Kickstarter. We analyzed over 8,000 Kickstarter projects and simulated decisions made by human backers, incorporating prompt engineering to test the consistency and fairness of AI-generated donation rationales and amounts. We explore whether LLMs are subject to the anchoring bias, similar to human subjects, and whether they are more likely to rely on human or machine anchors.

Results showed that the output varied with the prompt structure, even when based on identical project information. This finding highlights the importance of careful prompt engineering to ensure reliable AI-supported decision-making. The study's findings raise concerns about whether LLMs are susceptible to the same cognitive biases as people are. The broader goal is to understand LLM behavior in order to mitigate cognitive biases for subjective tasks.

3 - Designing Rigorous Experiments with Generative AI: Opportunities, Threats, and Best Practices**Weifeng Li, University of Georgia, Athens, GA, United States, Carolina Alves de Lima Salge**

Generative AI (GenAI) is transforming experimental research by enabling rapid, scalable, and adaptive workflows across all stages of the research process—from stimulus design and pilot testing to follow-up and analysis. This paper outlines how GenAI can be leveraged to improve efficiency and rigor while identifying emerging threats across Shadish et al.'s (2002) validity framework. I draw on recent applications in information systems and experimental economics to show how GenAI facilitates standardization, uniform treatment delivery, synthetic pretesting, and AI-assisted data analysis. However, I also identify risks such as construct ambiguity, instrumentation drift, and AI-induced bias, particularly when model behavior is uncontrolled or unvalidated. To navigate this evolving landscape, I develop a structured framework of threats and mitigation strategies, as well as eleven normative best practices. These practices span the full experimental lifecycle, including guidance on pre-registration, disclosure, output validation, and AI-human equivalence testing. I also highlight the importance of transparency, model documentation, and ethical safeguards to address privacy, autonomy, and reproducibility concerns. By synthesizing technical capabilities with methodological rigor, this work offers both a roadmap and a caution for researchers integrating GenAI into experimental design. The contribution is twofold: (1) a validity-based evaluation of GenAI's impact on experimental methods, and (2) actionable recommendations for responsibly deploying GenAI in complex, information-rich domains such as information systems research. This guidance is particularly relevant for scholars seeking to enhance scalability, reduce costs, and preserve causal inference while using GenAI-powered tools.

TA10

Building A Level 3 A312

Resource Allocation and Defender-attacker Problems

Job Market Showcase

Job Market Showcase

Chair: Samuel Affar, University of Tennessee, Knoxville, Knoxville, TN, United States

1 - Trustworthy Algorithmic Solutions: from Fair Division to Federated Learning**Aniket Murhekar, University of Illinois, Urbana-Champaign, Urbana, IL, United States**

In today's world, algorithmic decision-making plays a crucial role in shaping social and industrial outcomes. A fundamental challenge, however, is ensuring that these algorithmic solutions are trustworthy – that is, fair to the participants, economically efficient, and aligned with their incentives. A wealth of remarkable works address such

challenges in various settings, though they are largely focused on the distribution of rivalrous goods.

In this talk, I will present my work on designing trustworthy algorithms and solution concepts for non-traditional and modern settings such as task allocation and data collaboration. For the fair allocation of tasks, I will discuss our recent breakthrough showing the first constant-factor approximation of EFX, a compelling fairness criterion whose existence remains one of the most important open problems in fair division.

I will then turn to collaborative settings involving data such as federated learning and data exchange. I will present solution concepts that incentivize participation, discourage free-riding, and ensure fairness and welfare guarantees in these settings.

I will conclude with interesting avenues for future work in trustworthy resource allocation and data economics.

2 - Allocation and Pricing in Modern Settings: Submodularity, Online Arrivals and Data Markets

Pooja Kulkarni, UIUC, Urbana, IL, United States

The problem of allocating limited resources among competing agents arises across a range of operational contexts: assigning compute nodes to jobs, distributing bandwidth across network flows, allocating ad slots to bidders, or distributing aid among individuals. A central challenge in these settings is *fairness*—a criterion critical for ensuring system stability, social equity, and long-term participation.

My work addresses fair allocation under two non-standard but increasingly relevant settings:

- **Beyond-Additive Valuations:** Traditional models assume additive utilities, but real-world preferences often exhibit diminishing returns—captured by *submodular* and more general valuation functions. I will discuss challenges and results on fairness objectives such as **Maximin Share (MMS)** and **Nash Social Welfare (NSW)** under these richer valuation models.
- **Online Arrivals:** In many real-world systems, agents arrive over time and allocation decisions must be made without full information. These problems are notoriously hard with no previous positive results. I will talk about our new learning-augmented model for fair allocation in such *online* settings. Our model provides strong theoretical guarantees while still being able to capture real-world settings.

Closely related to allocation is the problem of *pricing*. Classical equilibrium models rely on the balance of supply and demand—a paradigm that breaks down for *data*, a replicable and non-rivalrous resource. In the second part of my talk, I will discuss our new notion of equilibrium for data markets and present its existence and other desirable properties for valuations arising naturally from *vertically partitioned datasets*.

3 - A Successive Refinement Algorithm for Tri-Level Stochastic Defender-Attacker Problem with Decision-Dependent Probabilities

Samuel Affar, University of Tennessee, Knoxville, Knoxville, TN, United States

Tri-level defender-attacker game models are a well-studied method for determining how best to protect a system (e.g., a transportation network) from attacks. Existing models assume that defender and attacker actions have a perfect effect, i.e., system components hardened by a defender cannot be destroyed by the attacker, and attacked components always fail. Because of these assumptions, these models produce solutions in which defended components are never attacked, which may not be realistic in some contexts. This study considers an imperfect defender-attacker problem in which defender decisions (e.g., hardening) and attacker decisions (e.g., interdiction) have an imperfect effect such that the probability distribution of a component's capacity depends on the amount of defense and attack resources allocated to the component. Thus, this problem is a stochastic optimization problem with decision-dependent probabilities and is challenging to solve because the deterministic equivalent formulation has many high-degree multilinear terms. To address the challenges in solving this problem, we propose a successive refinement algorithm that dynamically refines the support of the random variables as needed, leveraging the fact that a less-refined support has fewer scenarios and multilinear terms and is, therefore, easier to solve. A comparison of the successive refinement algorithm versus the deterministic equivalent formulation on a tri-level stochastic maximum flow problem indicates that the proposed method solves many more problem instances and is up to 66 times faster. These results indicate that it is now possible to solve tri-level problems with imperfect hardening and attacks.

4 - Defensive Resource Allocation Versus a Realistic Attacker in a Multi-Layer Security Game

Ian Unson, University at Buffalo, Buffalo, NY, United States, Jun Zhuang

Understanding how adversaries make decisions is essential for enhancing security and protecting public spaces. This project develops new tools to help security professionals allocate resources more effectively, minimizing vulnerabilities at soft targets such as schools, public venues, and places of worship. By combining insights from human-subject experiments and surveys with advanced mathematical modeling, this research provides practical strategies for making better defensive decisions. Unlike traditional, theoretical approaches, which assume adversaries act purely rationally, this project incorporates data from human decision-making experiments and surveys to create more realistic and actionable models. The outcomes of this research extend beyond improving security as they also offer valuable insights into adversarial behavior that can inform fields like marketing and risk management. By advancing our understanding of decision-making and optimizing resource allocation, this project promotes national safety and prosperity. This project develops a novel framework for studying adversarial decision-making by combining game-theoretic modeling with data from human-subject experiments and surveys. The primary objective is to model an attacker's best response function based on actual decisions made in attacker-defender scenarios, rather than assuming purely rational and theoretical strategies. The research investigates key variables, reward, cost, and probability of success and their influence on adversarial behavior. Using this data, the project derives closed-form Nash equilibria to optimize defensive resource allocation strategies across multi-layered security systems. The results contribute to actionable resource allocation strategies while offering a scalable framework for analyzing adversarial behavior in diverse applications, including security and risk management.

TA11

Building A Level 3 A313

Smart and Autonomous Systems

Job Market Showcase

Job Market Showcase

Chair: Sriparvathi Shaji Bhattachiri, Rochester Institute of Technology, Rochester, NY, 14623, United States

1 - Mutual Information Surprise: Rethinking Unexpectedness in Autonomous Systems

Yinsong Wang, Georgia Institute of Technology, Atlanta, GA, United States

Recent breakthroughs in autonomous experimentation have demonstrated remarkable physical capabilities, yet their cognitive control remains limited—often relying on static heuristics or classical optimization. A core limitation is the absence of a principled mechanism to detect and adapt to the unexpectedness. While traditional surprise measures—such as Shannon or Bayesian Surprise—offer momentary detection of deviation, they fail to capture whether a system is truly learning and adapting. In this work, we introduce Mutual Information Surprise (MIS), a new framework that redefines surprise not as anomaly detection, but as a signal of epistemic growth. MIS quantifies the impact of new observations on mutual information, enabling autonomous systems to reflect on their learning progression. We develop a statistical test sequence to detect meaningful shifts in estimated mutual information and propose a mutual information surprise reaction policy (MISRP) that dynamically governs system behavior through sampling adjustment and process forking. Empirical evaluations—on both synthetic domains and a dynamic pollution map estimation task—show that MISRP-governed strategies significantly outperform classical surprise-based approaches in stability, responsiveness, and predictive accuracy. By shifting surprise from reactive to reflective, MIS offers a path toward more self-aware and adaptive autonomous systems.

2 - Time-Series Analytics for Smart Manufacturing Systems

Mojtaba Farahani, University of South Carolina, Columbia, SC, United States, Thorsten Wuest

We are witnessing a transformation in manufacturing industries driven by the integration of Industry 4.0 infrastructure and smart manufacturing systems. Leveraging technologies like Machine Learning (ML), Artificial Intelligence (AI), and Digital Twins (DT) by manufacturers on the shop floor and using data-driven decision-making, leads to unlock efficiency gains, quality improvements, and cost savings. This work focuses on time-series analytics within smart manufacturing systems, aiming to bridge the existing research gap between theoretical advancements in computer science literature and their practical application in manufacturing with Time-Series Classification, and Time-Series Forecasting identified as the primary tasks. The main objective is to develop a comprehensive framework that will help to bridge the identified research gap, provide practical guidance for practitioners, advance smart manufacturing systems, and promote promising Research Avenues. This framework aims to advance the maturity of time-series analytics in manufacturing, providing valuable guidance for researchers and practitioners.

3 - Integrating AI and Simulation for Intelligent Material Handling Using Autonomous Mobile Robots

Sripavathi Shaji Bhattathiri, Rochester Institute of Tech., Rochester, NY, United States

This research tackles challenges in integrating autonomous mobile robots (AMRs) into material handling by addressing two interconnected problems using AI-driven solutions: intelligent dispatching and seamless human-robot interaction (HRI). Motivated by the need for safe, efficient automation, the work addresses dynamic task allocation and non-verbal communication in shared spaces.

First, a reinforcement learning (RL) framework optimizes dynamic task assignments, outperforming rule-based methods (e.g., shortest distance) under moderate deadline tightness by balancing travel distance and due dates. It matches rule-based efficiency under tight deadlines, adapts to various warehouse layouts, and scales in simulated environments.

Second, for HRI, experiments show that unidirectional intent signaling (e.g., digital avatars) improve coordination, but seamless HRI often requires bidirectional communication. A deep learning-based gesture recognition system using RGB cameras achieves $\geq 99\%$ accuracy, enabling AMRs to interpret human gestures without verbal cues. The robustness of the model is tested by synthetically altering the background as well. A gesture language framework and machine learning pipeline together contribute to robot perception paving way further to intuitive collaboration.

Together, these contributions advance AMR deployment in real-world dynamic environments: RL enhances task efficiency, while vision-based gesture recognition ensures safety. Results demonstrate AI's potential to harmonize productivity and safety in industrial automation, offering scalable solutions that combine theoretical advances with practical needs. This work paves the way for smarter, collaborative robotic systems in dynamic intralogistics environments.

TA12

Building A Level 3 A314

Advances in Healthcare and Energy

Job Market Showcase

Job Market Showcase

Chair: Jaachinma Okafor, University of Tennessee, Knoxville, TN, 37920, United States

Co-Chair: Yusi Wei, Drexel University, Philadelphia, PA, United States

1 - Modeling Shared Ownership to Reduce Cost Barriers for Distributed Storage

Ogechi Vivian Nwadiaru, University of Massachusetts, Amherst, MA, United States

Energy storage, especially distributed energy storage, can provide key operational flexibility services and support the participation of diverse actors particularly low-income consumers in the energy transition. However, incentivising the adoption of energy storage behind-the-meter remains an open challenge. Whilst distributed energy storage can help provide affordable, reliable electricity access to economically marginalized communities, investment in residential storage is only economically viable with value stacking. The challenge with value stacking lies in the complexity of monetizing multiple revenue streams, such as energy arbitrage, demand charge management, and grid services, which often requires access to market mechanisms, regulatory support, and technological expertise that marginalized communities may lack. In this work, we evaluate the potential of novel equity-driven energy storage ownership structures combined with time-based demand response incentives in increasing the benefits marginalized communities derive from the energy transition. We investigate the value of shared energy storage systems (SES) and quantify the monetary savings derived from various aggregation sizes under three distinct time of use structures.

2 - Trade-Offs Between Privacy and Utility in Machine Learning

Yusi Wei, Drexel University, Philadelphia, PA, United States

The secondary use of healthcare data is vital for research and clinical innovation, but it raises concerns about patient privacy. This study investigates how to balance privacy preservation and data utility in healthcare data sharing, considering the perspectives of both data providers and data users. Using a dataset of adult patients hospitalized between 2013 and 2015, we predict whether sepsis was present at admission or developed during the hospital stay. We identify subpopulations—such as older adults, frequently hospitalized patients, and racial minorities—that are especially vulnerable to privacy attacks due to their unique combinations of demographic and healthcare utilization attributes. These groups are also critical for machine learning (ML) model performance. We evaluate three anonymization methods—k-anonymity, the technique by Zheng et al., and our proposed model, the MO-OBAM—based on their ability to reduce re-identification risk while maintaining ML utility. Results show that k-anonymity offers limited protection. The methods of Zheng et al. and MO-OBAM provide stronger privacy safeguards, with MO-OBAM yielding the best utility outcomes: only a 2% change in precision and recall compared to the original dataset. This work provides actionable insights for healthcare organizations on how to share data responsibly. It highlights the need for anonymization methods that protect vulnerable populations without sacrificing the performance of data-driven models.

3 - Sustainable Operations - Sustainable Supply Chains & Energy Servicing

Jaachinma Okafor, University of Tennessee, Knoxville, TN, United States

his talk presents two research streams that address key challenges in sustainable operations: expanding renewable energy access and reducing supply chain emissions.

The first study explores a *servicizing* model to improve residential energy access in developing countries. This model involves a firm installing and maintaining a solar array-battery system in households in exchange for usage-based payments. This removes upfront cost barriers and shifts ownership and maintenance responsibilities to the provider. We develop an analytical model to determine the optimal pricing, solar capacity, and battery size, incorporating cost, demand variability, and adoption dynamics. A numerical study using solar and demand data from Nigerian households validates the model and highlights the profitability and reliability of shared systems, especially in areas reliant on gasoline generators.

The second study examines how buyers can incentivize suppliers to voluntarily reduce greenhouse gas (GHG) emissions. Using a game-theoretic model and a behavioral experiment, We compare two mechanisms: a fixed bonus and a wholesale price premium. While the analytical model favors the price premium, the experiment reveals that a fixed bonus more effectively motivates emission reductions, emphasizing the role of behavioral factors in supply chain contracting.

Together, these studies offer practical insights for designing sustainable business models and incentive mechanisms that advance environmental and social goals across energy systems and supply chains.

TA13

Building A Level 3 A315

Novel Problems in Estimation

Job Market Showcase

Job Market Showcase

Chair: Danielle Morey, University of Washington, Seattle, WA, United States

1 - Design-Based Weighted Regression Estimators for Conditional Spillover Effects**Fei Fang, Yale University, New Haven, CT, United States**, Edoardo Airoldi, Laura Forastiere

In a clustered interference setting, with networks collected within clusters and no interference between clusters, we introduce a general causal estimand for conditional spillover effects, offering flexible ways of integrating unit-to-unit spillover effects. In particular, we define spillover effects from the treatment received by one neighbor, averaged over the distribution of the cluster treatment, and conditional on the characteristics of the treated unit. Such definition enables to access the heterogeneity of a unit's spillover effect on their neighbors with respect to the unit's characteristics. Two weighted regression-based estimators are proposed: i) at the individual level, taking neighbors' averages either in the outcomes or in the treatments within weights; and ii) at the dyadic level, where the outcome of one unit is regressed on the treatment of each neighbor. When covariates driving the heterogeneity are categorical, we prove the equivalence of the two regression-based estimators to the non-parametric Hajek estimator. For continuous covariates, we demonstrate that both estimators consistently estimate the proposed estimands. Under a design-based perspective, we derive HAC variance estimators and establish the central limit theorem. Simulations are conducted to compare the performance of our estimators. Finally, we apply our methods to a randomized experiment conducted in Honduras to evaluate the spillover effect of a behavioral intervention.

2 - Degradation Signal Missing Data Imputation via Transformers**Madi Arabi, North Carolina State University, Raleigh, NC, United States**, Xiaolei Fang

Accurate prognostics for industrial assets rely heavily on the quality and completeness of degradation signals, yet real-world scenarios often suffer from sparse, fragmented, or missing condition-monitoring data due to sensor failures, cost constraints, or invasive measurement procedures. To address this challenge, we investigate transformer-based imputation techniques for multivariate degradation signals with the goal of improving time-to-failure (TTF) prediction. While traditional methods such as statistical approaches and matrix factorization often disregard temporal dependencies, transformer architectures are well-suited for capturing long-range relationships in sequential data through self-attention. We benchmark the performance of standard transformers and convolution-enhanced transformers for imputing missing values and predicting TTF, under varying degrees of signal sparsity. Our results highlight the strengths and limitations of each model, and ongoing work focuses on architectural modifications to optimize imputation quality for downstream prognostic tasks.

3 - Computationally Efficient Reductions Between Some Statistical Models**Mengqi Lou, Georgia Institute of Technology, Atlanta, GA, United States**, Guy Bresler, Ashwin Pananjady

We study the problem of approximately transforming a sample from a source statistical model to a sample from a target statistical model without knowing the parameters of the source model, and construct several computationally efficient such reductions between canonical statistical experiments. In particular, we provide computationally efficient procedures that approximately reduce uniform, Erlang, and Laplace location models to general target families. We illustrate our methodology by establishing nonasymptotic reductions between some canonical high-dimensional problems, spanning mixtures of experts, phase retrieval, and signal denoising. Notably, the reductions are structure-preserving and can accommodate missing data. We also point to a possible application in transforming one differentially private mechanism to another.

4 - Optimal Design in the Absence of Ground Truth: a Multi-Model Consistency Approach**Danielle Morey, University of Washington, Seattle, WA, United States**

There exist many real-world complex systems that can be represented via multiple models, but where no known ground truth exists. Traditional methods of multi-fidelity modeling and optimization often fall short in such cases. Motivated by applications in underwater communication networks and biomanufacturing, I propose a weighted comparison methodology that leverages consistency between models to optimize systems without relying on ground truth.

TA14

Building A Level 3 A316

Drone-Assisted Delivery and Logistics Optimization

Contributed Session

Contributed

Chair: Bibek Yadav, Florida A & M university, 1001 Ocala Road, D328, Tallahassee, FL, 32304, United States

1 - The Drone Routing Problem with Mixed Fleet of Electric and Conventional Trucks**Cagri Koc, Hacettepe University, Ankara, Turkey**, Gulcin Bektur, Cihan Tugrul Cicek

We introduce a variant of the drone routing problem with a mixed fleet by considering both internal combustion and electric trucks. In the problem, both truck types are equipped with a single drone. A drone may either depart from the starting depot along with its vehicle or directly from the starting depot to serve a relevant customer. Customers must receive service either from a drone or a truck. However, due to the drone's limited carrying capacity, some customers may not be eligible for drone service. The drone can be launched either directly from the starting depot or from a node along the truck's route to serve a relevant customer. After serving a customer, the drone returns either directly to the ending depot or to the truck at a node along the vehicle's route. Once the truck picks up the drone at the meeting point, it can continue its route from the corresponding node. The objective is to minimize the maximum tour completion time. We formulate the problem as a mixed-integer linear programming model and introduce valid inequalities to strengthen the formulation. We develop a metaheuristic to effectively solve the problem. We conduct extensive computational experiments on a set of generated instances, based on several data sets from the literature, to provide insights into the problem's characteristics. Our metaheuristic proves to be highly effective in solving the problem for large-scale instances.

2 - Drone-Truck Arc Routing Problems**Sung Hoon Chung, Binghamton University, Binghamton, NY, United States**, saif al-nimer

We address a Drone-Truck Arc Routing Problem (DT-ARP), for which a clustering algorithm is proposed to improve computational efficiency. This problem extends the rural postman problem by incorporating an aerial vehicle (drone) in addition to the ground vehicle. We also investigate the optimal combination of drone launch and recovery nodes to minimize the makespan. Numerical examples are provided to demonstrate the effectiveness of the proposed model and heuristic for the DT-ARP.

3 - Drone Scheduling and Mobile Pantry Placement for Enhanced Campus Food Distribution**Tamanna Kamal, North Carolina State University, Raleigh, NC, United States**, Ameya Abhay Mulay, Syed Md Kashan Naiyer, Leila Hajibabai

Ensuring equitable food access on university campuses presents a growing logistical challenge, especially with the rising rates of food insecurity. Traditional fixed-location food distribution systems often fall short in meeting the spatially and temporally dynamic needs of campus communities. This study proposes a data-driven decision support framework that integrates a maximal covering location model with an approximate dynamic programming (ADP) algorithm to address these challenges. The framework optimizes mobile pantry placement and schedules real-time drone deliveries across campus. The location model determines optimal pantry sites using ZIP-

code-level food insecurity data to maximize geographic coverage and support distributional equity. The ADP algorithm dynamically coordinates drone routing and pantry pickups by incorporating practical constraints such as limited drone capacity, time windows, and urban mobility barriers. The model is informed by real-world inventory and demographic data from Feed the Pack, the food pantry at North Carolina State University. Preliminary results show that the integrated approach reduces unmet demand and improves operational responsiveness compared to static delivery strategies. Sensitivity analysis reveals key trade-offs between equity and efficiency, providing decision-makers with actionable insights for resource allocation and logistical planning. This work contributes an adaptive framework for campus food distribution that aligns operational efficiency with humanitarian objectives. The methods are generalizable to other institutional or urban contexts where mobility and equity are central concerns.

Keywords—Approximate dynamic programming, drone delivery, mobile pantry, food insecurity, humanitarian logistics, decision support.

TA15

Building A Level 4 A401

Aerial Vehicle Routing in Challenging Conditions

Invited Session

Air Transportation Section

Co-Chair: Garrett Schueller, University of Texas at Austin, 3001 Cedar st, Austin, TX, 78705, United States

1 - Optimizing Multi-Team Agricultural Drone Deployment and Routing with Weather Aware Planning

Ethan Kolby, University of Michigan, Ann Arbor, MI, United States

Optimizing agricultural drone operations requires addressing unique complexities posed by dynamic weather conditions, particularly in multi-team deployments. Recent research on facility location problems (FLP), vehicle routing problems (VRP), and coverage VRP tailored to agricultural applications emphasizes how weather considerations impact drone utilization strategies. Key areas explored include wide-region survey drone deployments, where real-time weather dynamics, such as wind and precipitation, may temporarily restrict or diminish flight efficiency. Integrating weather-aware routing strategies enhances mission adaptability and robustness. Additionally, crop spraying operations illustrate how predictive weather modeling influences route selection and adjustments. Incorporating weather forecasts into optimization models has the potential to improve spray coverage efficacy. The importance of weather-informed optimization strategies in agricultural drone logistics is highlighted throughout.

2 - Last-Mile Delivery Optimization of Combining Public Transport and Drone Systems

Bipan Zou, Zhongnan University of Economics and Law, Wuhan, China, People's Republic of, Rui Zhang, Jiliu Li, Yuanqian Shi

This paper considers an innovative last-mile delivery system that combines public transport with drones. We study the problem of delivering a given set of parcels from warehouses to customers using the drones, in collaboration with several bus lines, with the objective of minimizing the sum of the completion time of all orders. We first formulate this problem in a decomposition framework with the master problem determines a schedule for each drone that consists of the routing and bus taking decisions, and the subproblem aims at identifying promising schedules to be added into the master problem. We design a state function driven branch-and-price-and-cut (BPC) algorithm that uses a novel dynamic programming (DP) approach to find the promising schedules. A dominance rule and three strategies are developed to speed up the DP those bring substantial improvements. We apply a variable fixing technique to eliminate unpromising schedules in advance and design a heuristic to help solving the pricing subproblem. Using both a random instance and a real dataset, we verify the efficiency of our algorithm by comparing with commercial solvers. Numerical results show that the drone service area can be significantly expanded, especially when the drone battery capacity is medium. Besides, introducing public buses can promote the sharing of drones among warehouses, that reduces the total delivery time by up to 15%. Finally, we analyze several select mechanisms of bus lines and show the potential benefits of slightly adjusting bus lines for warehouse proximity alignment.

3 - Performance Estimation of Aircraft Computer Vision Models Using Optimal Transport-Based Distributional Mismatch

Binshuai Wang, The George Washington University, Washington, DC, United States, Peng Wei

Ensuring the reliable deployment of deep learning models in real-world environments remains a fundamental challenge, particularly in safety-critical applications such as autonomous unmanned aerial system (UAS) landings. This paper investigates the relationship between model performance and distributional mismatch between training data and operational design domain (ODD) samples. We propose a framework that quantifies this mismatch using Wasserstein distances derived from optimal transport theory, applied to both raw image data and semantic embeddings from a pretrained CLIP (ViT-B/32) model. To address the scarcity of real ODD data collected by flight tests, we additionally leverage diffusion models to generate additional ODD scenarios, and evaluate their quality using Fréchet Inception Distance (FID) score. Experimental results demonstrate a strong correlation between the magnitude of distributional mismatch and model performance degradation. Moreover, our results demonstrate that models trained on datasets with lower distributional mismatch to the ODD domain, such as VisDrone, compared to COCO, achieve superior detection performance. These findings highlight the effectiveness of our framework as a practical and generalizable tool for estimating model reliability in complex real-world deployment scenarios, when trained on different training datasets.

4 - Flying Against the Wind: Smart Uav Stationing for Faster Disaster Relief

Sina Kazemdehbashi, Wayne State University, Detroit, MI, United States, Yanchao Liu, Boris Mordukhovich

Natural and human-made disasters can cause severe devastation and loss of life. In such situations, rapid collection of information about affected areas, damages, and victims is critical. However, in many disasters—such as earthquakes—communication infrastructure is often destroyed, limiting victims' ability to send distress signals and hampering rescue efforts.

Unmanned Aerial Vehicles (UAVs) can assist by collecting data and establishing temporary communication networks. However, real-world deployment is challenged by factors such as wind and heterogeneous UAV dynamics. To address this, we extend classical single-facility location problems by incorporating heterogeneous dynamic sets that represent varying UAV speeds. We generalize the Fermat-Torricelli and Sylvester problems to introduce the Sylvester-Fermat-Torricelli (SFT) problem, which captures complex factors such as wind influence, UAV heterogeneity, and back-and-forth motion within a unified framework. The SFT problem exhibits intrinsic nonsmoothness due to the min-max nature of its objective function. To overcome the resulting analytical complexities, we employ tools from convex analysis.

Experiments show that our approach can reduce operation time by up to 55%, improving the efficiency of UAV-based disaster response missions.

5 - Probabilistic Modeling of Lunar Terrain for Uncertainty-Aware Global Path Planning via Stochastic Programming

Garrett Schueller, University of Texas at Austin, Austin, TX, United States, John-Paul Clarke, Alessandro Pinto

Accurate global path planning is essential for the success of lunar rover missions, yet current methods are fundamentally constrained by uncertainty in orbital imagery. Remote sensing offers only partial, low-resolution views of the surface, rendering deterministic plans brittle in the face of unexpected hazards. To address this, we present a framework that integrates computer vision and stochastic programming to enable robust, uncertainty-aware path planning. A convolutional neural network is trained to predict traversability distributions from LROC imagery, generating probabilistic cost maps that model terrain risk. The model demonstrates strong performance across standard evaluation metrics. To improve realism, spatial correlations are incorporated into the map sampling process, yielding coherent terrain realizations. These samples serve as inputs to a two-stage stochastic branch-and-bound algorithm, which selects global paths that balance expected performance with downside risk. This framework offers a scalable and extensible solution for risk-aware autonomy, demonstrated on a lunar test case and applicable to long-range exploration of other planetary surfaces.

TA16

Building A Level 4 A402

Optimizing Product Supply Reliability With Advanced Risk Models

Invited Session

The Practice Section of Informs

Co-Chair: Hanzhang Liu, Johnson & Johnson, Denville, NJ, United States

1 - Enhancing Supply Chain Resilience and Agility Through Advanced Planning Parameter Optimization

Nasim Arbabzadeh, Johnson & Johnson, Bellevue, WA, United States, VISHAL VARMA

The reliability provided by healthcare supply chains to their patients and customers is strongly contingent upon a reliable flow of input raw materials from their diverse supplier base. Any delays in receipt of critical raw materials can pose serious risks to an otherwise smoothly running supply chain. Predicting delays or volatility in raw material supply lead times is therefore an important component of managing today's highly inter-twined healthcare supply chains. This work presents an advanced time series-based modeling framework to predict supplier lead times, the applications of those lead times into various business workflows such as right-sizing inventories and raising critical supply alerts to material resource planners. We provide examples of specially designed hierarchical time series models, inspired by proportional-integral-derivative (PID) control principles, which can handle predictions at different levels of the material and supplier hierarchy and how we track their performance using multiple error metrics. Following successful pilot results, the solution is scalable for deployment across multiple global manufacturing sites and SAP platforms. Performance tracking using multiple error metrics shows how such models can be deployed at large scales and drive substantial improvements in supply reliability, resilience and operational efficiency.

2 - Designing Resilient Networks for Cellular Therapies with Mathematical Programming Based Heuristics

Aric Johnson, Johnson & Johnson, Raritan, NJ, United States

Advanced cellular therapies are increasingly offering the "new frontier" against several high mortality illnesses such as cancer. Delivery of these therapies at scale requires the underlying therapy supply chains to control cell processing lead times within tight limits. This in turn requires the availability of all supporting raw materials while ensuring adequate availability of cell/raw material production capacities, timeliness of raw material procurements and production runs, and minimal wastage of materials related to obsolescence. These are complex objectives to meet especially in a highly dynamic environment of rapidly growing demand for such therapies. In this talk we describe mathematical programming based heuristics designed for supporting advanced cell therapy supply chains and their implementations using latest sematic layer technologies.

3 - Optimizing Supplier Selection under Tiered Pricing and Resilience Constraints

Brijesh Jakhania, Johnson & Johnson, Allentown, PA, United States

Organizations that procure critical materials face a recurring trade-off between minimizing cost and maintaining resilience. Supplier quotes are commonly tiered—unit prices decline only after specific volume thresholds—while policies often mandate multiple suppliers or minimum/maximum share limits. These interacting rules, combined with shifting demand and evolving quotes, make manual allocation error-prone and expensive. This work presents a year-by-year mixed-integer optimization framework that selects suppliers, activates exactly one price tier per chosen supplier, and allocates volumes to meet requirements at the lowest total cost subject to resilience and share constraints. The approach uses tidy, auditable inputs (requirements by year, supplier tier tables, policy rules) and produces transparent outputs (selected suppliers, active tiers, volumes, and total cost) that can be reviewed by finance and procurement and rerun quickly as conditions change. We illustrate typical recommendation patterns—concentration on cost-effective tiers with measured diversification when policies require it—and quantify the incremental “price of resilience.” The framework is practical to implement, fast to execute, and straightforward to govern, enabling organizations to capture tier-based savings, make informed trade-offs, and institutionalize repeatable, defensible sourcing decisions.

4 - A Decomposition-Based Heuristic for Optimizing Inventories in A Multi-Echelon Inventory System

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

This study proposes a decomposition-based heuristic for determining order-up-to level at each stage within a general multi-echelon system, accounting for stochastic lead times resulting from random stock-outs throughout the system. In a multi-echelon inventory system, upstream stages (nodes) are responsible for supplying items to downstream stages, fulfilling their demands whenever sufficient inventory is available. Each downstream stage manages its own inventory while determining order quantities based on the inventory levels of the upstream stage. This interdependency emphasizes a critical relationship—shortages at the upstream stages can significantly impact the ordering decisions of downstream stages. This study explores both the theoretical and numerical aspects of how order quantities at downstream stages are influenced by upstream inventory levels and how these decisions evolve over a given time horizon. This understanding will provide an essential building block for developing the proposed heuristic.

TA17

Building A Level 4 A403

AI, Data Science, and Optimization in Media: Methods and Applications

Invited Session

Practice Track

Chair: Xin Ma, Monks, Atlanta, GA, United States

1 - Reimagining Media Decisions: How AI and ML are Transforming Media Planning and Optimization

Xin Ma, Monks, Atlanta, GA, United States

AI and ML are revolutionizing media planning, testing, and optimization by delivering actionable insights, enhancing targeting precision, and boosting campaign performance. This ultimately enables advertisers and brands to achieve higher ROI and more effective advertising strategies. In media planning, we train models using historical media and performance data to understand how media investments influence business outcomes. By factoring in elements like channel interactions, adstock, and saturation effects, we can generate investment scenarios and optimize budget allocations and fighting across channels and tactics. In the realm of testing, we offer a data-driven geo-experimentation tool that enables clients to design and evaluate incrementality effectively. By leveraging first-party data, demographic, and socioeconomic data, and applying advanced AI and ML algorithms, the tool identifies key performance drivers. It then crafts well-balanced test/control splits and provides recommendations on budget and test windows to achieve statistical significance, tailored to the specific testing needs.

2 - Precision Planning: A Simulation Tool for Addressable TV Impact and Reach

Wes Chaar, Avant-Garde Strategy, Dallas, TX, United States

In an era of precision marketing and evolving viewer behaviors, Addressable TV offers a powerful alternative to traditional broadcast planning - but demonstrating its value at the planning stage remains a challenge. This session will introduce a next-generation simulation tool designed to model the reach and performance of Addressable TV campaigns under a wide range of “what-if” scenarios. Attendees will explore how the tool empowers advertisers, agencies, and media strategists to forecast audience reach, assess incremental value, and optimize campaign design based on key variables such as spend, targeting, and media mix. With features like dynamic scenario comparisons, saturation diagnostics, and budget-efficiency projections, the tool transforms complex planning into clear, data-driven decisions. This session will show how scenario simulation can accelerate Addressable TV adoption by delivering transparency, flexibility, and actionable insight at the moment decisions are made.

3 - AI Applications in Media Measurements: Audience, Content, and Revenue Optimization

Hanxi Bao, NBCUniversal, New York, NY, United States

The media and advertising industry is undergoing a fundamental transformation driven by AI technologies. This presentation explores a few examples on leveraging innovations to address critical challenges in audiences testing, content recognition, and revenue allocations.

With the help of LLM and cloud data platforms, audience profiling and segmentation can be performed using a mix of 1st party and 3rd party data. Machine learning algorithms analyze large datasets to identify audience patterns, predict future behavior, and create unified audience profiles for media planners. This enables highly targeted content strategies and personalized user experiences that drive higher engagements and better impacts.

AI can also help on upper funnel media measurements through reach and frequency measurements and optimizations. AI tools can be implemented in the workflow of audience deduplication process to help eliminating impression wastes and maximizing unique reach across different platforms.

AI-assisted A/B Testing moves beyond traditional experimentation by using machine learning to automatically generate test hypotheses, conduct dynamic multivariate testing, and provide predictive optimization. This enables publishers to optimize content performance faster and more effectively than manual testing approaches.

Implementation requires a phased approach from data engineering, modeling developments and deployments, to the result summary and product managements. This presentation summarizes the examples of AI-driven advanced analytics from a publisher perspective and suggest guidelines in media measurements.

TA18

Building A Level 4 A404

Optimization Techniques for Contested Logistics

Invited Session

Military and Security

Chair: Robert Curry, University of Arkansas, Fayetteville, AR, United States

1 - Optimizing Unmanned Resupply to Marine Littoral Regiments

Gary Lazzaro, United States Naval Academy, Annapolis, MD, United States

The future of warfare in the littoral regions of the world presents a unique challenge of delivering supplies to forward deployed forces in contested areas. Unmanned Surface Vehicles (USVs) offer a contemporary approach to resupplying Marine Littoral Regiments (MLR) with the materials needed for sustained operations within the enemy's weapons engagement zone. We present an unclassified scenario where MLRs require resupply in an archipelago island chain that is defended by a number of enemy patrol craft. We utilize search and detection theory to create a user-friendly Excel interface. This tool provides warfare commanders with information regarding the probability of detection for a library of friendly USVs. It also analyzes specifications of friendly USV options to best evade enemy USV barrier patrols for different battlespace scenarios. It aids in deciding how to best employ the chosen friendly USVs to effectively resupply. We also present an integer program that chooses an optimal combination of friendly USVs to maximize survivability of craft while minimizing the cost of lost USVs. Our work enables future teams to use our tool in a classified environment for real-world scenarios.

2 - Tri-Level Optimization for Building Supply Chain Resilience with Safety Stock Against a Worst-Case Disruption

Dan Kosmas, University of Massachusetts Lowell, Lowell, MA, United States, Carson Grose, Ozlem Ergun

Engineering resilient supply chains is crucial to ensuring critical commodities such as fuel, pharmaceuticals, and emergency response supplies are available when needed. A resilient supply chain should function even in the event of a disruption, which can come in the form of natural disasters, equipment failures, or attacks by nefarious actors. One way to build resilience in a supply chain is through the pre-allocation of inventory, or safety stock to carefully chosen locations in the distribution network. We develop a tri-level minimum cost flow interdiction model to improve the resiliency of a distribution network by optimally pre-allocating safety stock while accounting for disruptions. We numerically demonstrate how different measures of supply chain resiliency – such as safety stock allocation, arc density, and production capacity – impact the total costs of operating the distribution network and the network service level.

3 - Optimizing Escort Strategies for Chokepoint and Littoral Maritime Lanes Using Mixed-Integer Programming

Shantonu Mitra, University of Arkansas, Fayetteville, AR, United States, Robert Curry, Kelly Sullivan

Over the years, the variety and sophistication of attacks on merchant ships in the littoral & chokepoint sea lanes – such as attacks using the explosive unmanned surface vessels (USVs), unmanned aerial vehicles (UAVs), modern anti-ship ballistic and cruise missiles have increased. So, the aim of this research is to allocate limited military resources to protect maritime shipping vessels such as oil tankers or aircraft carriers that need to traverse a shipping lane. To solve this problem, we have used the two-stage stochastic mixed-integer programming (MIP) approach where the first stage involves decisions such as assigning escort ships & allocating weapons and the second stage focuses on threat engagement & threat neutralization. The results suggest that the expected survival value of the vessels get decreased when the available weapon units of the escort ships aren't capable enough of neutralizing the threats as well as the late engagement of the threats by the escorts. The proposed model balances between the resource allocation versus threat neutralization and multi-scenario robustness versus resource consumption.

4 - Military Logistics Planning for Expeditionary Warfare

Mathieu Dahan, Georgia Institute of Technology, Atlanta, GA, United States, Matthew Ford, Samuel Tan, Haden Boone, Peter Frazier, Yongjia Song, Huseyin Topaloglu

We consider the problem of designing military logistics operations in a contested environment. The objective is to route connectors to transport critical supplies (e.g., fuel, ammunition) from their stocking locations to demand locations by fixed deadlines set by a fire plan. We formulate this problem as a variant of the scheduled service network design problem developed for commercial logistics service providers. In our model, commodities may be sourced from various locations, and are transported by sea and air connectors with heterogeneous speeds and capacities. The resulting mixed-integer program determines a consolidation plan, i.e., paths for connectors and commodities on a time-space network, to minimize the delay in satisfying demands. To obtain good quality solutions, we design heuristics that exploit the problem's characteristics. We demonstrate our methodology on realistic training instances developed by the United States Marine Corps and provide military planners with insights regarding the design of logistics operations in contested environments.

TA19

Building A Level 4 A405

Theoretical Frontiers in Causal Estimation

Invited Session

Applied Probability Society

Chair: Vasilis Syrgkanis, TBA, TBA, CA, United States

Co-Chair: Jikai Jin, Stanford University, Stanford, CA, United States

Co-Chair: Jiyan Tan

1 - Synthetic Survival Control

Devavrat Shah, Massachusetts Institute of Technology, Cambridge, MA, United States, Xinyi Han

In this talk, we shall present a natural generalization of causal inference method, synthetic control, with roots in traditional panel data to enable evaluation of counterfactual outcome where outcome corresponds to survival function.

2 - Stable Causal Estimation with Transport Maps

Yixin Wang, University of Michigan, Ann Arbor, MI, United States

This work proposes a causal estimation algorithm for observational studies using Optimal Transport (OT) maps. Traditional importance weighting based estimator often suffer from high estimation variance, especially when strong overlap is lacking, e.g. with high-dimensional covariates or small sample sizes. In contrast, OT maps offer a more stable approach to addressing distribution shift. To this end, we introduce an OT-based algorithm to estimate average treatment effects. To handle partially overlapping support, we further extend the algorithm using linear interpolation maps. Additionally, we develop a distributional sensitivity analysis framework using Wasserstein Distributionally Robust Optimization (WDRO). This approach relaxes the conditional exchangeability assumption and provides partial identification bounds for the causal estimand.

This is joint work with Yidan Xu and Long Nguyen.

3 - Consistency of Neural Causal Partial Identification

Jiyuan Tan, Stanford University, Stanford, CA, United States

Recent progress in Neural Causal Models (NCMs) showcased how identification and partial identification of causal effects can be automatically carried out via training of neural generative models that respect the constraints encoded in a given causal graph [Xia et al, Balazadeh et al]. However, formal consistency of these methods has only been proven for the case of discrete variables or only for linear causal models. In this work, we prove the consistency of partial identification via NCMs in a general setting with both continuous and categorical variables. Further, our results highlight the impact of the design of the underlying neural network architecture in terms of depth and connectivity as well as the importance of applying Lipschitz regularization in the training phase. In particular, we provide a counterexample showing that without Lipschitz regularization this method may not be asymptotically consistent. Our results are enabled by new results on the approximability of Structural Causal Models (SCMs) via neural generative models, together with an analysis of the sample complexity of the resulting architectures and how that translates into an error in the constrained optimization problem that defines the partial identification bounds.

4 - Structure-Agnostic Optimality of Doubly Robust Learning

Jikai Jin, Stanford University, Stanford, CA, United States, Vasilis Syrgkanis

The design of efficient non-parametric estimators has long been a central problem in statistics. However, optimal estimators designed under assumed structural properties crucially rely on the validity of these assumptions, so that they are often cumbersome to deploy in practice. In this paper, we consider non-parametric estimation problems for which the parameter of interest depends on an unknown nuisance function.

Motivated by the wide adoption of non-parametric adaptive estimators, we consider a structure-agnostic framework, where one does not impose structural assumptions on the model and only assumes access to black-box nuisance estimates. Within this framework, we first prove the minimax optimality of doubly robust learning for estimating the average treatment effect, arguably one of the central parameters of interest in the causal inference literature. We then generalize this result by establishing the optimality of the popular first-order debiasing estimators under general conditions that are satisfied by a broad range of parameter estimation problems. Notably, we differentiate between two regimes in which double robustness can and cannot be achieved and in which first-order debiasing yields different error rates. Our result implies that first-order debiasing is simultaneously optimal in both regimes. We instantiate our lower bound by deriving optimal error rates that recover existing results, as well as for new parameters that are of interest in the literature, including the weighted average derivative, average policy effect, counterfactual effect of distribution shift and the expected derivative of conditional quantile.

5 - Regression Adjustments for Experimental Designs in Two-Sided Marketplaces

Timothy Sudijono, Stanford University, Stanford, CA, United States

Multiple randomization designs (MRDs) are a class of experimental designs used to handle interference in two-sided marketplaces. We investigate regression adjustment strategies for estimating total, spillover, and direct effects in MRDs. We derive minimum asymptotic variance estimators among a broad class of linearly adjusted estimators. Surprisingly, the optimal regression adjustments are estimable from data and are generally different from regression adjustments in classical randomized experiments. For example, one such optimal estimator for the direct effect corresponds to a weighted regression with interacted two-way fixed effects. We establish model-robustness properties, central limit theorems, and inferential methods for our estimators; we also contribute to the theory of MRD experiments. These results provide the analog of the CUPED regression adjustment for marketplace experiments. Numerical simulations demonstrate a considerable increase in efficiency over simpler approaches, enabling better inference when running MRDs.

TA20

Building A Level 4 A406

Transportation Network Optimization

Invited Session

Railway Applications

Chair: Krishna Jha, Optym, Coppell, TX, 75019, United States

1 - Train Movement Optimization

Krishna Jha, Optym, Coppell, TX, United States, Wei Huang, Sajeeb Kumar Kirtonia

Optimizing train movements is a complex decision problem due to physical movement constraints, cascading decision impacts, and conflicting operational objectives. We have developed a solution approach that leverages advanced optimization and machine learning techniques to address this problem effectively. Our approach is scalable for large railroad networks and has been applied across multiple use cases, including improvements in service quality, network throughput, and maintenance planning. In this talk, we will present the methodology and share numerical results that demonstrate its effectiveness.

2 - Rail-Centric Multi-Objective Network Optimization for Sustainable Logistics of Mineral Tailings in Construction Supply Chains

Claudio Cunha, University of Sao Paulo, Sao Paulo, Brazil

This study addresses the challenge of designing a sustainable logistics network for the large-scale distribution of iron ore tailings, as by-product of mining operations in Brazil. When stored in dams, these tailings pose serious environmental risks—but they also offer a valuable opportunity for reuse as substitutes for raw materials in civil construction. However, their economic viability is strongly dependent on an optimized logistics system due to their low commercial value and the extensive distances to potential markets. To address this, we develop a rail-centric, multi-objective linear optimization model that integrates economic, environmental, and social factors. The model allowed us to determine optimal multimodal flows—primarily via rail but including road and waterway segments—linking a single origin to over 100 mesoregions through a capacity-constrained hub network. Objectives include maximizing total profit, minimizing negative socio-environmental impacts and maximizing positive externalities (e.g., job creation, GDP contribution). Using a constraint- ϵ method, we generate Pareto-optimal tradeoffs that allow policymakers and logistics planners to assess how marginal profit reductions can yield significant environmental and social benefits. Results indicate that a partial substitution strategy, guided by targeted regional selection and mode choice, enables a feasible, profit-generating network configuration that significantly reduces externalities. The framework underscores the strategic role of rail infrastructure in decarbonizing bulk logistics and demonstrates how optimized multimodal transport can link mining residues to sustainable construction supply chains. By enabling the integration of industrial by-products into the circular economy, it contributes to the decarbonization of both the construction and mining sectors through network optimization.

3 - Optimizing Maintenance Intervals Using Monte Carlo Simulation

Masood Kang, TTX Company, Charlotte, NC, United States, David Cannon, Clayton Bartelt

Condition-Based Maintenance (CBM) is a proactive approach scheduling maintenance based on equipment conditions rather than fixed intervals. In the context of freight rail rolling stock, CBM leverages sensors and diagnostic data to predict and prevent failures, thereby enhancing fleet reliability and availability. This method contrasts with traditional maintenance practices, which often lead to either premature or delayed overhauls. A critical aspect of CBM is determining the optimal End-of-Life (EOL) thresholds for components such as bolsters. EOL thresholds are the points at which components are deemed no longer reliable and require replacement. Setting these thresholds accurately is crucial for minimizing maintenance costs and avoiding unexpected failures. To address this, we developed a Monte Carlo Simulation model to evaluate different maintenance practices and identify the optimal intervals for EOL thresholds. The simulation incorporates various factors, including component failure probabilities, maintenance costs, and fleet health metrics. By iterating over numerous scenarios, the model provides insights into the impact of different EOL thresholds on overall fleet performance. Our findings indicate that adjusting EOL thresholds can significantly reduce maintenance costs and improve fleet reliability. For instance, increasing the EOL mileage thresholds for bolsters can lead to substantial cost savings while maintaining acceptable levels of reliability. The simulation results also highlight the importance of balancing maintenance capacity and work order generation to manage backlog effectively. This project demonstrates the potential of Monte Carlo Simulation in optimizing maintenance strategies for freight rail rolling stock, offering a data-driven approach to enhance CBM practices and achieve better operational outcomes.

4 - A Backtracking-Based Heuristic for Train to Fuel Pad Assignments

Ashutosh Shukla, BNSF Railway, Fort Worth, TX, United States, Dasaradh Mallampati, Paul Kuhn

Belen, N.M. is BNSF Railway's largest fueling station, annually serving close to 50,000 trains with fuel worth more than a billion dollars. These trains can have different costs when it comes to delays. That is, the delay of certain kinds of trains is preferred over others. At the same time, to make an optimal assignment of trains to fuel pads, one must also account for the network topology, train's fueling, crew change, and inspection time. Up until now, these assignments have been done manually. To automate this assignment process, we present a backtracking-based heuristic that adheres to all the hard and soft constraints and generates valid assignments within seconds using the most up-to-date train arrival status.

TA21

Building A Level 4 A407

Market Design and Network Games (II)

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Song Gao, University of Massachusetts Amherst, Amherst, MA, United States

Co-Chair: Kenan Zhang, EPFL, Lausanne, Switzerland

1 - A Markovian Traffic Equilibrium Model for Ride-Hailing

Song Gao, University of Massachusetts Amherst, Amherst, MA, United States, Chiwei Yan, Guocheng Jiang, Hanyu Cheng

We develop a Markovian traffic equilibrium model for ride-hailing, where ride-hailing vehicles make sequential link choices in a network, whether empty or hired, to maximize random utilities in an infinite-horizon semi-Markovian process. Competitions among empty vehicles for passengers are modeled at nodes, and traffic congestion due to road usage is modeled on links. We define the equilibrium as the solution to a fixed point problem, and prove the existence of the fixed point solution. We conduct computational experiments in networks with varying sizes to explore the convergence of the fixed point iteration, and derive policy and managerial insights. We contribute to the state of the art by developing a new equilibrium model that accounts for both traffic congestion and vehicle competition for passengers in a sequential decision making framework to account for drivers' forward looking capabilities.

2 - Mobility-as-a-Service (MaaS) System as a Multi-Leader-Multi-Follower Game: a Single-Level Variational Inequality (Vi) Formulation

Kenan Zhang, EPFL, Lausanne, Switzerland

This study models a Mobility-as-a-Service (MaaS) system as a multi-leader-multi-follower game that captures the complex interactions among the MaaS platform, service operators, and travelers. We consider a cooperative setting where the MaaS platform purchases service capacity from service operators and sells multi-modal trips to travelers following an origin-destination-based pricing scheme; meanwhile, service operators use their remaining capacities to serve single-modal trips. As followers, travelers make both mode choices, including whether to use MaaS, and route choices in the multi-modal transportation network subject to congestion. Inspired by dual formulation for traffic assignment problems, we propose a novel single-level variational inequality (VI) formulation by introducing a virtual traffic operator, along with the MaaS platform and multiple service operators. To decouple the decision-making process between the MaaS platform and service operators, we define additional variables, proposed capacity purchase price by the MaaS platform, and impose consensus constraints that also represent the negotiation process. A key advantage of the proposed VI formulation is that it supports parallel solution procedures and thus enables large-scale applications.

3 - Cooperation and the Design of Public Goods

Juan Martinez Mori, University of Colorado Denver, Denver, CO, United States, Alejandro Toriello

We consider the cooperative elements that arise in the design of public goods, such as transportation policies and infrastructure. These involve a variety of stakeholders: governments, businesses, advocates, and users. Their eventual deployment is critically dependent on the decision maker's ability to garner sufficient support from each of these groups; we formalize these strategic requirements from the perspective of cooperative game theory. Specifically, we introduce non-transferable utility, linear production (NTU LP) games, which combine the game-theoretic tensions inherent in public decision-making with the modeling flexibility of linear programming. We derive structural properties regarding the non-emptiness, representability, and complexity of the core, a solution concept that models the viability of cooperation. In particular, we provide fairly general sufficient conditions under which the core of an NTU LP game is guaranteed to be non-empty, prove that determining membership in the core is co-NP-complete, and develop a cutting plane algorithm to optimize various social welfare objectives subject to core membership. Lastly, we apply these results in a data-driven case study on service plan optimization for the Chicago bus system. We illustrate how, while cooperation is necessary for the successful deployment of transportation service plans, it may also have adverse and/or counterintuitive distributive implications.

4 - Data-Driven Facility Layout for Enhancing Walk Appeal in Walkable Neighborhoods

Haoran Guo, Tsinghua University, Beijing, China, People's Republic of, Sentao Miao, Wei Qi

In response to the challenges posed by car-centric urban development, there is growing global advocacy for walkable neighborhoods that promote public health and urban sustainability. To support the design of such neighborhoods, this paper studies the optimal layout of facilities (e.g., retail and service establishments) to enhance walk appeal. Unlike conventional designs that rely primarily on distance metrics, our approach focuses on pedestrian choices. We use a latent class logit model to estimate how facility layouts influence individuals' propensity to walk, accounting for heterogeneous travel purposes. This model is calibrated with a city-wide dataset from our industry collaborator, a leading Chinese map service provider. Building on this model, we develop a choice-based optimization framework to determine facility layouts that maximize the aggregate walking probability of residents. We first analyze its structural properties to derive analytical insights into optimal layout patterns. To find optimal layouts for this NP-hard problem, we reformulate it as an exponential cone program. To address the computational challenges this poses for large-scale applications, we further develop efficient greedy algorithms and establish their performance guarantees via submodularity ratio analysis. A case study in Beijing shows that our approach significantly improves walk appeal compared to traditional methods, with minimal impact on facility accessibility. Our results also reveal a "wisdom of crowds" effect in existing layouts and emphasize the importance of neighborhood context in shaping optimal design patterns. Specifically, we find that dispersed layouts are preferable in already walkable environments, while more challenging walking conditions benefit from spatially concentrated facilities.

TA22

Building A Level 4 A408

Experimentation and Interference

Invited Session

Applied Probability Society

Chair: Andrew Zheng, University of British Columbia, Boston, MA, United States

Co-Chair: Anushka Murthy, Stanford University, Stanford, CA, United States

Co-Chair: Wenqian Xing, Stanford University, Stanford, CA, United States

1 - Differences-in-Neighbors for Network Interference in Experiments

Andrew Zheng, University of British Columbia, Boston, MA, United States, Tianyi Peng, Naimeng Ye

Experiments in online platforms frequently suffer from network interference, in which a treatment applied to a given unit affects outcomes for other units connected via the platform. This SUTVA violation biases naive approaches to experiment design and estimation. A common solution is to reduce interference by clustering connected units, and randomizing treatments at the cluster level, typically followed by estimation using one of two extremes: either a simple difference-in-means (DM) estimator, which ignores remaining interference; or an unbiased Horvitz-Thompson (HT) estimator, which eliminates interference at great cost in variance. Even combined with clustered designs, this presents a limited set of achievable bias variance tradeoffs. We propose a new estimator, dubbed Differences-in-Neighbors (DN), designed explicitly to mitigate network interference. Compared to DM estimators, DN achieves bias second order in the magnitude of the interference effect, while its variance is exponentially smaller than that of HT estimators. When combined with clustered designs, DN offers improved bias-variance tradeoffs not achievable by existing approaches. Empirical evaluations on a large-scale social network and a city-level ride-sharing simulator demonstrate the superior performance of DN in experiments at practical scale.

2 - Experimentation Under Non-Stationary Interference

Su Jia, Cornell, New York, NY, United States, Peter Frazier, Nathan Kallus, Christina Yu

We consider experimentation in the presence of non-stationarity, inter-unit (spatial) interference, and carry-over effects (temporal interference), where we wish to estimate the global average treatment effect (GATE), the difference between average outcomes having exposed all units at all times to treatment or to control. Most work in this area assume that the interference between units is encoded as a *static* graph. However, in many real world settings, this interference graph varies over time, for example, since the units are moving. We study the scenario where the interference graph changes in a deterministic fashion. As the main result, we show that a natural design and estimator achieves an MSE of $1/NT$ times a factor that measures the complexity of the system, where N and T are the number of units and time periods.

3 - Estimation of Treatment Effects Under Nonstationarity via Truncated Difference-in-Q&R's

Ramesh Johari, Stanford University, Stanford, CA, United States, Tianyi Peng, Wenqian Xing

Randomized controlled experiments ("A/B tests") are essential for evaluating interventions in dynamic data-driven systems, where actions affect both current and future states. Accurate estimation of the global average treatment effect (GATE) must consider nonstationarity in these temporal dynamics; unfortunately, existing estimators typically suffer from high bias, high variance, or both in nonstationary environments. We propose the truncated DQ estimator, extending difference-in-means to use truncated outcome trajectories. We prove it estimates a policy gradient approximating the GATE and derive bias and variance bounds. Simulations and empirically calibrated synthetic experiments demonstrate its low bias and variance under nonstationarity.

4 - Heterogeneous Treatment Effects in Matching Marketplaces Under Interference: from Estimation to Decision

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

Marketplace platforms frequently employ randomized experiments to guide operational decisions, such as determining whether to implement a new intervention globally or for specific user groups. Understanding how treatment effects vary across subpopulations—known as heterogeneous treatment effect (HTE) analysis—is critical for these decisions. However, such experiments often involve interference: the treatment assigned to one user can influence the outcomes of others. We first show that under interference, HTEs may not be uniquely defined, posing a fundamental challenge to identification and estimation. When the HTE for a subgroup is uniquely defined, we demonstrate that an estimator derived from linear programming duality can systematically reduce bias compared to the standard estimator. In settings where HTEs are not uniquely identifiable, we reframe the problem: rather than estimating effects directly, we focus on determining whether to implement the intervention within each subgroup. We leverage the duality-based estimator to construct a robust decision rule that accounts for interference and yields improved performance over rules derived from standard estimators, offering a better balance between effectiveness and conservatism. Theoretical findings are supported by numerical experiments on synthetic data, validating the proposed methodology.

TA23

Building A Level 4 A409

Applications of Simulation

Contributed Session

Contributed

Chair: Fatemeh Mahroo, Clemson University, 811 Issaqueena Trl, Clemson, 29630, United States

1 - Modeling Life Settlements

Joseph Bakhtiar, Georgia Tech, Atlanta, GA, United States, Dave Goldsman, John-Paul Clarke

We develop a comprehensive discrete-time framework for the valuation, risk management, and performance analysis of life settlement portfolios. After establishing a discrete-time mortality model employing various probability distributions calibrated with empirical data via maximum likelihood estimation, individual policy valuations incorporate analytical solutions, and Monte Carlo (MC) simulations are used to assess key financial metrics, including expected policy value over time, variance, and the probability of profitability. Additionally, we explore portfolio-level valuation by aggregating multiple life settlement contracts, explicitly accounting for correlated mortality risks. Through detailed MC simulations, we illustrate how mortality correlations impact portfolio diversification, expected returns, and risk measures such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). Robustness analyses evaluate the effects of model misspecifications, particularly concerning mortality assumptions and interest rate environments. Our discrete-time framework provides critical insights for investment decisions, portfolio optimization, and regulatory capital planning in the growing life settlements market.

2 - Proposal of the Shaft Theory: Analyzing Corporate Distress Using Concepts from Mechanics of Materials

Kaneki Yamaguchi, Waseda University, Tokyo, Japan

In today's volatile business environment, firms constantly face dramatic shifts in external and competitive conditions. Even companies with strong financial performance can rapidly collapse, as exemplified by the 2023 case of Bigmotor company in Japan. In response to increasing scrutiny, Japan mandated the disclosure of non-financial information, including human capital data, starting in 2023. To ensure sustainable growth, companies must now accurately assess their own value while monitoring competitors.

Although several corporate evaluation models attempt to assess a firm's sustainability and bankruptcy risk using both financial and non-financial metrics, few have focused on managerial resources—particularly human capital. This research addresses this gap by proposing a new corporate evaluation framework based on a physical analogy from materials mechanics: the "Shaft Theory." Specifically, the study models the structure of a firm's human resources using the mechanical properties of a shaft, one of the most fundamental components in structural mechanics.

Using analytical techniques from strength of materials, the model quantifies the "strength" of a firm's human resource structure and tests its relationship to business resilience and likelihood of survival. By establishing a statistically significant correlation between mechanical strength indicators and real-world corporate durability, the

research aims to validate the conceptual and practical value of Shaft Theory. This interdisciplinary approach not only provides a novel lens to understand organizational stability but also contributes to a more holistic evaluation framework that integrates human capital as a core factor in long-term corporate viability.

3 - Modeling Airport Tnc EV Charging Demand and Charging Infrastructure Needs Using Discrete-Event Simulation

Yanbo Ge, National Renewable Energy Laboratory, Golden, CO, United States, Monte Lunacek, Yi He, Kenneth Kelly

Transportation Network Companies (TNCs), such as Uber and Lyft, have set ambitious goals for fleet electrification. Airports are well-positioned to support TNC electric vehicle (EV) charging due to high passenger volumes, longer average trip distances, and naturally extended vehicle dwell times while waiting for ride requests. Accurately understanding infrastructure needs and charging demand is critical for both airports and TNCs as they plan for this transition. However, there is a lack of insight within both industry and academia regarding the energy requirements needed to support large-scale TNC electrification. To address this gap, the Athena ZEV project—funded by the U.S. Department of Energy's Vehicle Technologies Office (VTO)—has developed a discrete-event simulation model to quantify TNC EV charging infrastructure needs and energy demand at airports. The model tracks vehicle status and ride request dynamics at each time step, enabling scenario analysis based on varying assumptions such as vehicle state-of-charge (SOC), charger type and quantity, and fleet composition. A case study of Dallas/Fort Worth International Airport demonstrates the model's capabilities, highlighting how SOC policies, solar and storage integration, and charger design affect power demand and service quality. The tool can be adapted to any airport given the appropriate input data, offering a flexible and practical solution for planning TNC electrification strategies.

4 - Inventory Management Under Disruptions Using Risk-Averse Deep Reinforcement Learning

Chiye Ma, Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, China, People's Republic of, Wei Wang

Modern supply chains have become increasingly vulnerable and inefficient due to a range of disruptions, including trade wars, pandemics such as COVID-19, natural disasters, cyberattacks, and geopolitical conflicts. These challenges underscore the urgent need for effective inventory management strategies that can enhance supply chain resilience and ensure operational continuity. In this work, we propose a risk-averse multi-agent proximal policy optimization (MAPPO) framework to address inventory management problems under typical disruptive events. We focus on a two-echelon supply chain network consisting of one supplier and four retailers. The periodic review inventory control problem is formulated as a decentralized partially observable Markov decision process (Dec-POMDP), extended with a disruption context space to capture environmental uncertainties. To tackle the high-risk nature of such scenarios, we integrate MAPPO with the cross-entropy soft-risk (CeSoR) algorithm, enabling risk-aware policy optimization for the retailers. Numerical experiments demonstrate that our proposed method outperforms risk-neutral MAPPO under disrupted scenarios. The results highlight the effectiveness of our approach in improving inventory performance and enhancing supply chain resilience.

5 - Modeling EV Adoption Under Behavioral Heterogeneity: a Simulation Framework Using LLMs and Agent-Based Modeling

Xinyi Du, Singapore Management University, Singapore, Singapore, Junfeng Jiao

Achieving Singapore's Green Plan 2030 target for full electric vehicle (EV) adoption requires more than infrastructure; it depends on how residents respond to policy design. This study evaluates the behavioral impact of a default EV enrollment policy that automatically assigns new car buyers to electric vehicle pathways, combined with modest financial incentives. We develop a simulation framework that integrates large language models (LLMs) with behavioral economics to generate a heterogeneous synthetic population. LLM-guided narrative generation, grounded in microeconomic theory, enables us to move beyond fixed trait distributions by deriving agent-level characteristics, such as risk aversion, time discounting, and institutional trust, that reflect realistic behavioral biases and socio-demographic variation. These traits inform decision rules within an agent-based model (ABM) that simulates adoption behavior under the default-plus-incentive mechanism. We experiment with variations in default strength, opt-out friction, and incentive magnitude. Agent-level thresholds for adoption are shaped by behavioral parameters such as perceived risk and policy trust, allowing us to simulate differentiated response dynamics across population segments. The simulation also allows policymakers to test targeted scenarios, such as tailoring incentives to high-friction groups, to evaluate efficiency-equity trade-offs in a controlled setting. This study offers a virtual policy lab for testing and refining EV strategies prior to real-world implementation.

6 - A Simulation-Optimization Framework for Inference of Mixture Patience Distributions in Multi-Server Queues

Fatemeh Mahroo, Clemson University, Clemson, SC, United States, Tugce Isik

We present an iterative simulation-optimization framework for estimating customer patience in multi-server queueing systems with unobservable customer types and possible abandonment. Our goal is to accurately estimate patience parameters while minimizing server usage cost, customer abandonment cost, and costs associated with adding or removing servers. Our approach involves data generated by two parallel systems: the “real” system, where patience parameters are unknown, and a simulated “estimated” system, where these parameters are updated over time. The two systems operate under identical arrival rates, service rates, and service policies. At each iteration, the waiting times, abandonments, and service completions in the real system are observed, and the equivalent data is collected from the estimated system via simulation. We compare the waiting time distributions for served and abandoned customers using the Anderson–Darling two-sample test and the Wasserstein distance. The estimated system's parameters are updated using the Expectation-Maximization algorithm, and staffing levels are optimized to minimize operational costs. The algorithm terminates when the two statistical measures suggest that the two systems' patience behavior is indistinguishable. Computational experiments show that the method typically converges to the true patience parameters, with faster convergence observed when more abandonment occurs. The Wasserstein criterion is generally satisfied earlier than Anderson–Darling. By efficiently learning the underlying patience parameters, the method brings operational costs in line with those of a system with full information.

TA24

Building A Level 4 A410

Large-Scale Convex Optimization: Algorithms & Analyses via Monotone Operators

Tutorial session

Tutorials

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

1 - Large-Scale Convex Optimization: Algorithms & Analyses via Monotone Operators

Ernest Ryu, UCLA, Los Angeles, CA, United States, Wotao Yin

This tutorial presents a unified analysis of convex optimization algorithms through

the abstraction of monotone operators. Through this streamlined approach, we derive

and analyze a wide variety of classical and modern algorithms, including: Gradient

Descent, Dual Ascent, Proximal Point Method, Proximal Gradient Method, Projected

Gradient Method, Forward-Backward Splitting, Peaceman–Rachford Splitting,

Douglas–Rachford Splitting, Davis–Yin Splitting, Method of Multipliers, Proximal

Method of Multipliers, Alternating Direction Method of Multipliers, Alternating Minimization Algorithm, Primal-dual hybrid gradient (PDHG), PDL, and Condat–Vũ.

TA25

Building A Level 4 A411

Freight Management

Invited Session

TSL: Freight Transportation

Chair: Justin Goodson, Saint Louis University, Saint Louis, MO, United States

1 - Data-Driven Freight Shipment Reservation Problem

Raviv Pichayavet, Georgia Institute of Technology, Atlanta, GA, United States, Haoyun Wang, Yunqin Zhu, Alan Erera, Yao Xie, Benoit Montreuil

This project addresses the challenges faced by companies that rely on outsourced carriers to ship goods to their customer: securing cost-effective, reliable capacity for on-time pickups. Without their own delivery fleets, these companies depend on contracted carriers to ensure competitive shipping rates for specific lanes. However, these contracts lack guaranteed capacity, as carriers serve multiple shippers and allocate capacity on a first-come, first-served basis. When capacity is fully utilized, shippers face two costly choices: sourcing higher-priced services from alternative carriers or the spot market, or risking delays that impact operational schedules and customer satisfaction. To reduce the likelihood of last-minute costly purchases, the shipper has to make advance capacity reservations. However, this strategy is complicated by demand uncertainty, as shippers must reserve space for orders that may not yet be finalized. This study develops a data-driven stochastic optimization model to establish an effective reservation decision, addressing uncertainties in carrier capacity and shipment demand. By leveraging historical data on demand patterns, carrier capacity, and rate fluctuations, the model prescribes daily reservation quantities based on available capacity. It is designed to adapt as new information on demand and capacity becomes available, allowing the reservation policy to dynamically respond to changing conditions. This approach minimizes expected shipping costs while reducing the risk of capacity shortages and delays in scheduled pickups, providing companies with a flexible and responsive framework for managing outsourced shipping effectively.

2 - Spatio-Temporal Pattern Mining and Optimization for Load Consolidation in Freight Transportation Networks

Sikai Cheng, AI4OPT - Georgia Tech, Atlanta, GA, United States, Amira Hijazi, Jeren Konak, Alan Erera, Pascal Van Hentenryck

Freight consolidation is important to reduce transportation costs and mitigate congestion and pollution. An effective load consolidation plan relies on carefully chosen consolidation points to ensure alignment with existing transportation management processes, such as driver scheduling, personnel planning, and terminal operations. This complexity represents a significant challenge when searching for optimal consolidation strategies. Traditional optimization-based methods provide exact solutions, but their computational complexity makes them impractical for large-scale instances and they fail to leverage historical data. Machine learning-based approaches address these issues but often ignore operational constraints, leading to infeasible consolidation plans. This work proposes SPOT, the first approach that integrates the benefits of machine learning (ML) and optimization for load consolidation. The ML module plays a key role in the planning phase by defining the consolidation points through spatiotemporal clustering and frequent itemset mining, while the optimization module selects the most cost-effective feasible consolidation routes for a given operational day. Extensive experiments conducted on industrial load data demonstrate that SPOT significantly reduces travel distance and transportation costs compared to the existing industry-standard load planning strategy and a strong neighborhood-based heuristic. In addition, the ML module provides valuable tactical-level insights by identifying frequently recurring consolidation opportunities that guide proactive planning. Notably, SPOT is computationally efficient and can be easily scaled to accommodate large transportation networks.

3 - Dynamic Truck Scheduling Using Estimated Times of Arrival

Justin Goodson, Saint Louis University, Saint Louis, MO, United States

The paper addresses the problem of dynamically scheduling inbound trucks at a warehouse. Real-time vehicle positioning technology facilitates rolling estimates of truck arrival times. By taking these estimates as noisy signals of uncertain arrival times, we formulate a partially-observed Markov decision process to minimize expected delay in the assignment of trucks to unloading docks. A lookahead policy posts strong performance relative to a dual bound, significantly improves on the extant literature, substantially outperforms current practice, and scales to an industrial setting. Notably, using information relaxations and an information penalty, our analysis shows how mixed integer programming can be used to establish a lower limit on the cost of an optimal policy. The paper makes practical and methodological contributions to an industry with thin margins and demanding lead time targets.

4 - Line Network Design with Handling TIMES

Gabriel Deza, Tel Aviv University, Tel Aviv, Israel, Michal Tzur, Tal Raviv

We study the design of line-based vehicle routes for urban parcel delivery systems that rely on service points (SPs) and fixed periodic vehicle routes. These Hyperconnected Networks (HCNs) support transshipment between SPs, allowing multi-leg routing to boost consolidation and vehicle utilization. However, a key operational feature often overlooked in existing models is the significant time couriers spend idle at SPs handling parcels. Empirical data show that the majority of a driver's operational cycle is spent curbside—locating, scanning, loading, and unloading parcels—yet most optimization models treat this time as negligible or exogenous.

We introduce the Line Network Design with Handling Times (LND-HT) problem, which jointly optimizes line selection, parcel routing, and vehicle scheduling while explicitly accounting for parcel handling operations. We propose several formulations based on multi-commodity flow models over an expanded, multi-line network representation. We leverage the combinatorial problem structure to devise a branch-and-price solution method and compare with off the shelf solvers. Through numerical experiments on synthetic and real-world instances, we quantify the role of handling times in line planning for urban parcel delivery.

TA26

Building A Level 4 A412

Recent Advances in Ranking and Selection

Invited Session

Simulation

Chair: Eunhye Song, Georgia Tech

Co-Chair: Jaime Gonzalez-Hodar, Georgia Institute of Technology, Atlanta, GA, United States

1 - Data-Driven Ranking and Selection with Simultaneous Input Data Collection and Simulation

Enlu Zhou, Georgia Institute of Technology, Atlanta, GA, United States, Yuhao Wang

We propose a general and novel formulation of ranking and selection with the existence of streaming input data. The collection of multiple streams of such data may consume different types of resources, and hence can be conducted simultaneously. To utilize the streaming input data, we aggregate simulation outputs generated under heterogeneous input distributions over time to form a performance estimator. By characterizing the asymptotic behavior of the performance estimators, we formulate two optimization problems to optimally allocate budgets for collecting input data and running simulations. We then develop a multi-stage simultaneous budget allocation procedure and provide its statistical guarantees such as consistency and asymptotic normality. We conduct several numerical studies to demonstrate the competitive performance of the proposed procedure.

2 - Feature Misspecification in Sequential Learning Problems

Dongwook Shin, HKUST Business School, Clear Water Bay, Hong Kong, Dohyun Ahn, Assaf Zeevi

We consider a class of sequential learning problems where a decision maker must learn the unknown statistical characteristics of a finite set of alternatives (or systems) using sequential sampling to ultimately select a subset of “good” alternatives. A salient feature of our problem is that system performance is governed by a set of features. The decision maker postulates the dependence on these features to be linear, but this model may not precisely represent the true underlying system structure. We show that this mis-specification, if not managed properly, can lead to suboptimal performance because of a phenomenon identified as sample-selection endogeneity. We propose a prospective sampling principle—a new approach that eliminates the adverse effects of misspecification as the number of samples grows large. The proposed principle

applies across a very general class of widely used sampling policies, enjoys strong asymptotic performance guarantees, and exhibits effective finite-sample performance in numerical experiments.

3 - Ranking and Selection with Multiple Correct Answers

David Eckman, Texas A&M University, College Station, TX, United States, Raahim Hashmi, Taeho Kim

We consider ranking-and-selection problems in which the decision maker would be equally satisfied with selecting any of a subset of systems satisfying some generic definition of acceptability. Practical examples include epsilon-optimal or top-m systems in the single-objective setting or Pareto-optimal systems in the multi-objective setting. Problems of this kind have been studied by both the operations research and computer science communities under different assumptions and limiting regimes, which has resulted in fundamentally different algorithm designs. More specifically, operations researchers have at times adopted a large-deviations perspective that focuses on the rate at which the probability of making an incorrect selection approaches zero as the sampling budget increases. In contrast, computer scientists have more often focused on the worst-case sampling complexity needed to make a correct selection with some fixed confidence. We compare and contrast these paradigms in an effort to unify and deepen understanding about this important class of problems.

4 - General-Purpose Ranking and Selection for Stochastic Simulation with Streaming Input Data

Jaime Gonzalez-Hodar, Georgia Institute of Technology, Atlanta, GA, United States, Eunhye Song

We study ranking and selection (R&S) where the simulator's input models are increasingly more precisely estimated from the streaming data obtained from the system. The goal is to decide when to stop updating the model and return the estimated optimum with a probability of good selection (PGS) guarantee. We extend the general-purpose R&S procedure by Lee and Nelson by integrating a metamodel that represents the input uncertainty effect on the simulation output performance measure. The algorithm stops when the estimated PGS is no less than $1-\alpha$ accounting for both prediction error in the metamodel and input uncertainty. We then propose an alternative procedure that terminates significantly earlier while still providing the same (approximate) PGS guarantee by allowing the performance measures of inferior solutions to be estimated with lower precision than those of good solutions. Both algorithms can accommodate nonparametric input models and/or performance measures other than the means (e.g., quantiles).

TA27

Building B Level 2 B201

GPU Acceleration of Optimization Algorithms II

Invited Session

Computing Society

Chair: Matthew Galati, Amazon, Glen Mills, PA, United States

1 - Accelerating Crossovers with GPUs

Samiran Kawtikwar, University of Illinois Urbana Champaign, Champaign, IL, United States, Rakesh Nagi

This project explores the GPU-accelerated implementation of a novel crossover algorithm for linear programming (LP). The algorithm leverages the spiral dynamics observed in the primal-dual hybrid gradient (PDHG) method to efficiently transition from an interior or non-vertex optimal solution to a vertex optimal solution without resorting to traditional simplex-based methods. Our approach capitalizes on the inherent Single Instruction Multi Data (SIMD) parallelism of the PDL solver, specifically employing simultaneous perturbations of cost coefficients to enable efficient primal and dual pushes. This parallel strategy significantly enhances the suitability of the crossover process for GPU acceleration. Initial analysis highlights the substantial parallel potential of this approach, promising significant computational performance improvements. Ongoing computational experiments aim to quantify these performance gains and provide concrete speedup metrics. This work thus demonstrates a promising new direction for enhancing crossover processes in LP through GPU optimization.

2 - Second-Order Gpu Power for Machine Learning and Stats Procedures

Yan Xu, SAS, Cary, NC, United States, Joshua Griffin, Riadh Omhenni

A first-order method like SGD and coordinate-descent are attractive algorithms in that they are easier to implement and tend to work well in practice with nominal tuning. By design, they operate using vector-based (as opposed to matrix-based) operations. It is well-known that GPUs offer the greatest advantages for deep learning models that rely on fast dense-matrix-matrix multiplies. A challenge for first-order methods is that dense-matrix-matrix products do not naturally occur, and something like excessive batch sizes must be artificially induced. A benefit of second-order methods is that they are globally convergent and have naturally occurring dense matrix-matrix operations that are both cache-efficient on modern CPUs and naturally benefit from GPU devices. We will show the results of second-order methods applied to support vector machines and quantile regression, showing the additional benefit of GPU devices that customers may have purchased initially for applications in deep learning.

3 - How to Leverage a Gpu to Improve Mip/Lp Performance?

Mitch McKay, SimpleRose, St. Louis, MO, United States

In this talk we will discuss how designing algorithms for CPU vs GPU affect performance. With CPUs hitting the limit of Moore's law, we need to look to other hardware and novel algorithms to solve the ever growing problem of complexity in optimization. Recent hardware advances from Nvidia buck this trend, and enable a paradigm shift in how MIPs and LPs are solved. We will provide a developer's perspective on what we have learned at SimpleRose from implementing MIP/LP algorithms on NVIDIA GPUs.

4 - Training from Structure, Not from Data: Dataless Neural Networks for Combinatorial Optimization

Alvaro Velasquez, University of Colorado Boulder, Boulder, CO, United States, Cunxi Yu, Ismail Alkhouri

Recent advancements in deep learning have fueled a surge in research on machine learning for combinatorial optimization. These methods are typically rooted in (un)supervised learning and reinforcement learning. However, these approaches rely on labeled datasets, exhibit weak generalization, and often depend on problem-specific heuristics. In this presentation, we make the argument for dataless neural networks that are trained based on the structure of the problem instance as opposed to being trained on a conventional dataset. In particular, we present ReLU-based and quadratic relaxations for the MIS problem. By employing a gradient-based optimization algorithm like ADAM and leveraging efficient off-the-shelf GPU implementation, our straightforward yet effective approaches demonstrate competitive or superior performance compared to state-of-the-art learning-based methods.

TA28

Building B Level 2 B202

Federated Learning and Optimization II

Invited Session

Computing Society

Chair: Farzad Yousefian, Rutgers University, Piscataway, NJ, United States

Co-Chair: Kibaek Kim, Argonne National Laboratory, Lemont, IL, United States

1 - On Iteratively Regularized First-Order Methods for Simple Bilevel Optimization

Sepideh Samadigharabalah, Rutgers university, Piscataway, NJ, United States, Daniel Burbano, Farzad Yousefian

We consider simple bilevel optimization problems where the goal is to compute among the optimal solutions of a composite convex optimization problem, one that minimizes a secondary objective function. Our main contribution is threefold. (i) When the upper-level objective is a composite strongly convex function, we propose an iteratively regularized proximal gradient method in that the regularization parameter is updated at each iteration under a prescribed rule. We establish the asymptotic convergence of the generated iterate to the unique optimal solution. Further, we derive simultaneous sublinear convergence rates for suitably defined infeasibility and suboptimality error metrics. When the optimal solution set of the lower-level problem admits a weak sharp minimality condition, utilizing a constant regularization parameter, we show that this method achieves simultaneous linear convergence rates. (ii) For addressing the setting in (i), we also propose a regularized accelerated proximal gradient method. We derive quadratically decaying sublinear convergence rates for both infeasibility and suboptimality error metrics. When weak sharp minimality holds, a linear convergence rate with an improved dependence on the condition number is achieved. (iii) When the upper-level objective is a smooth nonconvex function, we propose an inexact projected iteratively regularized gradient method. Under suitable assumptions, we derive new convergence rate statements for computing a stationary point of the simple bilevel problem. We present preliminary numerical experiments for resolving three instances of ill-posed linear inverse problems.

2 - An Incentive-Enabled Method with Complexity Guarantees for Federated Learning

Fateme Maleki, Rutgers University, Piscataway, NJ, United States, Farzad Yousefian

Federated learning enables distributed model training without sharing raw data. In this framework, keeping the clients engaged remains challenging since participants may receive uneven benefits from their contributions. We consider nonconvex and L -smooth local objective functions for the clients. We present an Incentive-enabled Federated Averaging method that rewards participants based on their level of participation in each round, through allowing clients to control the size of their training data dynamically in each round of communication. In each communication round, clients exchange both local parameters with the server and their updated training dataset sizes, which are dynamically adjusted using a Nash equilibrium seeking update rule. Subsequently, the server transmits updated global parameters back to clients to initiate the following round. We theoretically analyze the proposed method, obtaining explicit iteration and communication complexity guarantees. We plan to conduct numerical experiments on MNIST and CIFAR-10 datasets to evaluate the effectiveness of our approach, with a focus on examining the convergence behavior of the loss function.

3 - Iteratively Regularized Gradient Tracking Methods for Distributed Optimal Equilibrium Seeking

Farzad Yousefian, Rutgers University, Piscataway, NJ, United States

We consider a class of distributed constrained optimization problems where the constraint set is characterized by the solution set of a distributed monotone variational inequality problem. This problem is motivated by the need for estimation of the efficiency of equilibria in Nash games. First, we consider solving this problem over directed networks. We develop an iteratively regularized distributed gradient tracking method where the agents employ a push-pull protocol to communicate over the network. Second, we consider a stochastic variant of this problem over undirected networks and develop an iteratively regularized distributed stochastic gradient tracking method. For both algorithms, we establish the convergence of the generated iterates by the agents to the optimal equilibrium and derive new convergence rate statements. We validate the two proposed methods and present preliminary numerical results for computing the optimal equilibrium in a Cournot competition.

TA29

Building B Level 2 B203

Advancing Hydrogen Integration in Energy Systems: Models and Analyses of Technical, Economic, and Legal Dimensions

Invited Session

ENRE: Energy

Chair: Chiara Lo Prete, The Pennsylvania State University John and Willie Leone Family Department of Energy and Mineral Engi, University Park, PA, United States

1 - Green Hydrogen Integration to Enhance Grid Reliability During Winter Emergencies

Jiaxing Wu, The Pennsylvania State University, State College, PA, United States, Chiara Lo Prete, Enzo Sauma

Increasing reliance on natural gas for electricity generation has heightened the power grid's vulnerability during extreme weather events, when gas demand spikes across both the electricity and heating sectors. This dependence on a single fuel source exposes the gas-fired generators to gas disruption risks and, consequently, challenges grid reliability. Green hydrogen, produced via water electrolysis powered by renewable energy, is emerging as a promising alternative, and could serve as a backup fuel during gas supply disruptions. This study investigates whether the integration of green hydrogen can improve power system reliability during winter emergencies by using electrolyzers to produce hydrogen when renewable energy is abundant, storing it on-site, and supplying it to gas-fired generators during natural gas shortages. This integration presents potential operational trade-offs: electrolyzers connected to the grid can increase electricity demand, potentially worsening transmission congestion—especially when they are located far from renewable energy sources and require power to be transmitted across already constrained parts of the grid. To assess potential benefits and challenges, this study couples a gas-electric optimization model with green hydrogen production and use. The integrated model quantifies the trade-off between enhanced reliability during natural gas disruptions due to scheduling reductions and increased grid congestion caused by the electricity demand of electrolyzers during non-critical periods. It captures the interactions among fuel availability, power system operations, and transmission constraints associated with green hydrogen integration.

2 - Unlocking the Value of Hydrogen Energy Storage: Modeling, Tools, and Assessment Insights

Di Wu, Pacific Northwest National Laboratory, Richland, WA, United States

Hydrogen is a promising solution for future energy systems. Understanding how to optimally utilize hydrogen energy storage (HES) to unlock stacked value streams is essential for advancing its market viability. This presentation introduces a techno-economic assessment framework for modeling hydrogen production, compression, storage, and utilization, capturing system interactions, and optimizing economic benefits across multiple value streams. We will also showcase the Hydrogen Energy Storage Evaluation Tool (HESET), a web-based application built on this framework. Finally, insights from real-world assessments of hydrogen and hybrid systems will be shared.

3 - Hydrogen in the Pipeline: Legal and Regulatory Challenges of Converting Natural Gas Pipelines for Hydrogen-Gas Blends

Chiara Lo Prete, Pennsylvania State University, University Park, PA, United States, Shahin Mirzaei

Hydrogen is the most abundant element in the universe, and its versatility and potential to lower emissions make it central for decarbonization purposes. As interest in hydrogen continues to grow in the United States, one proposed strategy for scaling up its use is blending small quantities hydrogen with natural gas in existing pipelines and, eventually, converting these pipelines to be able to carry higher hydrogen-natural gas ratios. However, the legal and regulatory frameworks governing the use of pipelines for hydrogen-gas blends remain underdeveloped. This presentation aims to examine the major legal challenges associated with pipeline repurposing for hydrogen-gas blends. Key topics include legal challenges associated with approving pipeline abandonment or conversion and addressing impacts on existing shippers' contracts. The goal of this presentation is to inform audience about the legal impediments to converting gas pipelines to carry gas-hydrogen blends and to shed light on the regulatory and legal barriers that must be addressed to enable the safe and reliable usage of hydrogen in the U.S. energy system.

4 - Co-Optimizing Freight and Hydrogen Logistics for Regional Decarbonization

Mohammad Reza Ghorbanali Zadegan, University of Texas at Austin, Austin, TX, United States, Zhaomiao Guo

Decarbonizing transportation and industrial operations has become a high priority as climate concerns intensify. Hydrogen plays a central role in national decarbonization plans and serves as a flexible alternative fuel in transport and industry, e.g., powering fuel-cell electric vehicles (FCEVs) or serving as feedstock for ammonia synthesis. However, large-scale hydrogen adoption requires coordinated planning of production and fueling infrastructure while addressing economic, technical, policy, and environmental challenges. This research develops a location-allocation optimization model to guide the deployment of hydrogen infrastructure for the heavy-duty

transportation sector and industrial applications. Decision variables cover station siting, electrolyzer sizing, hydrogen shipments, and freight refueling assignments, while the objective minimizes combined infrastructure, production, and transportation costs. This model provides insight into the evolution of infrastructure configurations and cost structures under different scenarios, and it serves as a guide to policymakers and industry in their quest for timely and economic decarbonization integrating hydrogen.

TA30

Building B Level 2 B204

Decarbonizing Energy Systems under Various Sources of Uncertainties

Invited Session

ENRE: Electricity

Chair: Angela Flores, Universidad de Chile, Santiago, Chile

Co-Chair: Rodrigo Moreno, University of Chile, Santiago, Chile

1 - Transmission Expansion Planning Under Weather-Induced Uncertainties

Angela Flores, Universidad de Chile, Santiago, Chile

Power systems are increasingly exposed to weather-induced uncertainties, driven by climate change. In Chile, rising temperatures, prolonged droughts, and more frequent extreme weather events such as heatwaves and storms pose growing risks to the power system operation. In this study, we develop a comprehensive transmission expansion planning model that accounts for both short term and long term climate-related uncertainties through multi-scale stochastic programming. We also present a methodology to generate multistage scenario trees that represent future weather-driven variability, using historical meteorological data in combination with downscaled climate projections. The proposed framework is applied to the Chilean power system and provides insights into how climate uncertainty may influence future transmission investment needs.

2 - Tractable Formulations for Mixed-Integer Integrated Medium- and Low-Voltage Distribution Expansion Planning with Controllable Devices

Alfredo Oneto, ETH Zurich, Zurich, Switzerland, Blazhe Gjorgiev, Giovanni Sansavini

The global transition to net-zero societies transforms power distribution networks (DNs), driven by the widespread adoption of distributed energy resources (DERs). These resources include, but are not limited to, photovoltaic systems, battery energy storage systems, electric vehicles, and heat pumps. While these technologies introduce new flexibility capabilities, they also increase volatility and strain networks initially designed for passive demand. Ensuring reliable and cost-effective operation under these evolving conditions requires accurate modeling of DER behavior and flexibility, as well as appropriate planning of DNs. This work contributes to that effort by planning integrated medium- and low-voltage systems with controllable DERs and applying complexity reduction techniques to maintain tractability.

The distribution infrastructure of Switzerland is analyzed as a case study, focusing on 2030, 2040, and 2050. Given the significant computational burden of analyzing the entire system, spatial and temporal aggregation techniques are applied. Specifically, previous results on clustering DNs using distributional and vectorial features to identify representative systems are used to reduce model construction time. In addition, a temporal aggregation method describes weekly seasonal storage with a reduced set of operational days. Furthermore, the investment decisions involve many integer variables, resulting in a challenging combinatorial problem. Suitable solution techniques, incorporating additional cuts and binary expansions, are applied to solve the optimization problem.

3 - The GB and Australian Path to Net Zero: How Do Regulators Navigate Epistemic Uncertainty, Multiple Criteria, and Diverse Stakeholders?

Evangelia Spyrou, Imperial College London, London, United Kingdom

Majority of the countries have net zero targets for future energy systems. Among the countries with detailed decarbonization plans and accountability frameworks, Great Britain and Australia are the largest greenhouse gas emitters and they both have ambitious plans to reach deep decarbonization levels in the near future. To achieve their ambitious goals, they have already changed several of their power system operating practices. For every change, regulators have to consider multiple criteria, diverse perspectives from stakeholders, and potential evolution of technological uncertainty. This presentation will summarize findings from the review of hundreds of changes adopted by the GB and Australian regulators in the last decade. The findings will highlight what are common trade-offs among criteria, how epistemic uncertainty affects the transformation of the sector and its pace, and how stakeholder participation varies among initiatives and types of stakeholders.

4 - An EU-Wide Electricity Generation Capacity Expansion Problem Subject to Network Constraints

Anthony Papavasiliou, National Technical University of Athens, Kifisia, Greece, Daniel Felipe Avila, Marilena Zampara, Georg Thomassen, David Pozo

The European Union has set ambitious renewable energy integration targets over the following decades as part of its climate, environmental protection and sustainability policy. These targets translate into concrete minimum targets for the supply of renewable energy sources, in particular wind and solar energy. Our focus in this work is to formulate and solve an EU-wide power generation expansion problem with such renewable energy targets. The model that we formulate provides an optimal generation expansion plan while accounting for uncertainty in future European climate scenarios as well as constraints in the ability of the power network to transport electric power from locations of abundant renewable supply to locations of high electricity demand.

TA31

Building B Level 2 B205

Planning Under Uncertainty and Cross-Sector Optimization

Contributed Session

Contributed

Chair: Stefano Galelli, Cornell University, Ithaca, NY, United States

1 - Priority List Heuristic for the Stochastic Unit Commitment Problem

Luis Rojas, Universidad Católica del Norte, Antofagasta, Chile, Carlos Olivares

The Unit Commitment Problem (UCP) is a fundamental optimization challenge in power systems operations, aiming to determine the optimal schedule of generating units to meet demand while minimizing costs. Incorporating renewable energy introduces stochastic generation patterns that significantly increase the computational burden. To address this, we propose a priority list heuristic for the Stochastic UCP, which generates an initial solution while ensuring that technical unit commitment constraints are met. This significantly reduces the computational time when searching for integer solutions. This starting point enables faster convergence of the full stochastic optimization procedure while preserving solution quality. Results from a case study using different power system instances and algorithmic details will be presented.

2 - Trends and Directions in Healthcare Supply Chain Research: a Data Mining Approach

Seunghye Ha, School of Business Administration, Kyungpook National University, Daegu, Korea, Republic of, Sungsu Kim

This study analyzes research trends in the field of healthcare supply chain management and suggests future research directions by applying topic modeling to approximately 500 international academic papers published between 2001 and 2024, retrieved from the Scopus and Web of Science databases. The analysis identified eight major research topics, including "Demand and Delivery Optimization," "Healthcare Logistics and Inventory Management," and "Crisis Response and Resource Allocation during Pandemics." In recent years, emerging topics such as "Supplier Selection and Risk Mitigation" and "Blockchain and Technology Applications" have gained increasing attention. The study also identifies central keywords such as Technology, Hospital, Procurement, Blockchain, and Supplier, which frequently appear in relation to major themes.

This study provides a systematic trend analysis based on international literature, offering insights into the thematic development and recent trajectories of research in the field of healthcare supply chains. By presenting representative papers for each topic, the study serves as a practical reference for researchers seeking to define actionable research questions and explore diverse applications. Furthermore, by identifying future research directions through quantitative analysis, the study aims to contribute to the academic accumulation and theoretical advancement of healthcare supply chain research.

3 - Optimizing Grid Planning Under Climate Uncertainty

Pin-Chun Cho, The University of Texas at Austin, Austin, TX, United States, Erhan Kutanoglu

Climate change is impacting the power grid in multiple ways compelling to rethink the processes and models used for grid planning. Rising temperatures, changing wind patterns, and evolving load profiles, and the variability of such factors across different locations are introducing new challenges to the spatial allocation of grid assets. In particular, identifying where to locate new renewable capacity and which existing plants to retire or reinforce has become more complex under climate uncertainty. We present a multi-objective optimization framework to support location-specific decisions for renewable generation siting, capacity replacement, and transmission reinforcement, incorporating high-resolution climate projections and operational constraints. The underlying model aims to identify infrastructure planning strategies that remain effective across a range of climate-shifted scenarios. We simulate weather-dependent load and renewable generation profiles using future projected climate data, capturing regional variation in future grid stress and resource availability. Using the optimization framework based on stochastic programming, we explore trade-offs among costs, emissions, and reliability in long-term planning. By explicitly modeling spatial and temporal correlations and climate-related uncertainty, we provide insights into how planning strategies should be optimized over time across multiple regions. A case study with a test grid on the footprint of Texas will show the potential benefits of optimization-based decision models linked to most up-to-date climate data to improve siting and retirement decisions, addressing a few of the challenges in energy transition under climate uncertainty.

4 - Assessing Temporal Aggregation Error in Capacity Expansion Models with Integrated Renewables, Grid, and Storage Constraints

Daniel Fontecha, KAUST, Jeddah, Saudi Arabia, Ricardo M. Lima, Omar Knio

Long-term investment decisions in the power sector require robust capacity expansion models (CEM) that minimize costs while meeting policy and reliability goals. Climate data, an essential input for CEM with renewables, is widely available, but its large size is cumbersome for linear optimization models with polynomial complexity. To address this, time aggregation techniques are used to reduce the dataset size, at the expense of accuracy. A well-designed aggregation method must therefore balance computational tractability and model fidelity.

A-priori clustering techniques, which define representative days before system design, are widely used but are inherently myopic, often leading to performance gaps when validated against the full dataset. More recent a-posteriori aggregation methods iteratively refine data selection using feedback from optimization results and have shown superior accuracy for models without spatial or temporal binding constraints. However, such constraints (like ramping limits, storage dynamics, and transmission flows) are fundamental to modern CEM and critical to capturing the operational flexibility of renewables.

This study evaluates the performance of two a-priori and one a-posteriori temporal aggregation techniques within a CEM with renewable and storage technologies, ramping limits, and inter-nodal transmission. We apply these methods to two case studies: a fully renewable system and a scenario with a fixed 2050 emission target. Using over 30 different climate datasets, we ensure statistical significance. The aggregation technique performance is assessed in terms of system cost, lost load, emissions, and computation time. Based on our findings, we offer practical recommendations and propose an enhanced aggregation approach informed by observed limitations.

5 - Coupled Modeling of Hydropower and Grid Operations in Regional Power Markets

Stefano Galelli, Cornell University, Ithaca, NY, United States, Hisham Eldardiry, Phumthep Bunnak

As wind and solar power increasingly shape electricity markets, hydropower is expected to play a more prominent role in maintaining grid flexibility and reliability. This shift underscores the need to understand how hydropower dispatch decisions respond to changing hydrological and demand conditions. Yet modeling these decisions remains difficult, as they depend simultaneously on the evolving states of both river basins and power systems. Traditional modeling soft-coupling approaches fall short, as they typically enforce one-way information flow, failing to capture dynamic feedbacks between water and energy systems.

Here, we present a tightly integrated modeling framework that hard-couples a reservoir system model (VIC-Res) with a large-scale power system simulator (PowNet). VIC-Res represents and optimizes hydropower operations, while PowNet performs unit commitment and economic dispatch. A dedicated coupler coordinates both models, synchronizing their time steps and ensuring convergence in simulated hydropower generation before advancing the system state.

We apply this framework to the Lao PDR–Thailand–Malaysia–Singapore Power Integration Project (LTMS-PIP), which depends heavily on hydropower from the Mekong and Chao Phraya basins. Our setup includes four PowNet instances (one per national grid) and two VIC-Res models (one per basin). Over a decade-long simulation, we demonstrate that integrating the hydrological and electrical system states into dispatch decisions improves overall system performance—lowering operational costs and reducing CO₂ emissions.

TA32

Building B Level 2 B206

Navigating Deep Climate Uncertainty in Energy Modeling

Invited Session

ENRE: Energy-Climate

Chair: Martha Christino, University of Michigan, Ann Arbor, MI, 48109, United States

Co-Chair: Michael Craig, University of Michigan, Ann Arbor, MI, United States

1 - Hydrochips: Coupler-Ready High Performance Pliers for Integrating River-Reservoir and Power System Models in Hydropower Management

Sean Turner, Oak Ridge National Laboratory, Oak Ridge, TN, United States, Stefano Galelli, Quentin Ploussard, Matija Pavicevic, Phumthep Bunnak

Hydropower operators face a daily challenge of coordinating power generation across their assets while faced with uncertain inflow and electricity price conditions. Performing this task may involve first optimizing generation using Unit Commitment / Economic Dispatch tools, then testing the resulting schedule with river routing and reservoir models to ensure feasibility with respect to river and reservoir dynamics. This is often a manual, multi-departmental process that can be inefficient and suboptimal, potentially leading to spilled energy and missed generating opportunities. We introduce the concept of a Coupler-ready High-Performance Plier (CHiPP) for addressing water-energy problems of this nature. CHiPPs are fast, open architecture tools, able to address the key dynamics of power and river systems models applied in the hydropower industry while being amenable to external control via a coupler. Development of CHiPPs will allow for in-depth experimentation and exploration into the value of hard-coupled water-power models. Using the Cumberland River System as an example case, we will present “Powersheds”—a high performance CHiPP for hourly water operations simulation in multi-reservoir cascades. Powersheds converts a desired multi-asset generating schedule to a sequence of water releases, considering turbine efficiencies with varying pool and tailwater elevations, reservoir bathymetry, river routing and other nuances of the river system. By identifying infeasibilities in the desired generating schedule, Powersheds is primed for interacting with production cost models by providing automated feedback on river system constraints. This research advances operations research methodologies for hydropower management.

2 - Combining Synthetic Generators, Machine Learning, and Open-Source Dcopf Models for Current and Future Grid Stress-Testing

Jordan Kern, North Carolina State University, Raleigh, NC, United States, Duc-Huy Pham, Troy Wibowo

Compound extremes (i.e., simultaneous occurrence of extreme values across regions and/or variable types) may pose the greatest resource adequacy risks for power systems. Given the scales and the number of variables involved, for any assumed state (past, present, or future), available weather datasets inevitably fail to capture all events that could pose risks to system reliability. Trying to address this limitation by increasing the sample size risks selection of data from non-representative system (i.e., climate) states. To overcome this, synthetic weather generators can be trained on empirical data and then used to generate new time series of unlimited size. Here, we demonstrate a framework for creating synthetic weather time series for use in predictive machine learning models of electricity load, wind power production, and solar power production, aggregated to the balancing authority (BA) level. We apply a heuristic dimensionality reduction to gridded weather data and then train a synthetic weather generator on this reduced data. We create synthetic ensembles of weather data across spatial scales ranging from a single electrical balancing authority (BA) to the contiguous U.S., and then pass these data through ML models trained to predict hourly electricity load, wind, and solar capacity factors for every BA in the U.S. We show how the resultant synthetic datasets can be used to perform stress testing of bulk power systems using open-source nodal DCOPF models developed in house. The “full stack” (synthetic weather generator, ML models, and grid simulation models) is publicly available for the entire contiguous U.S.

3 - Understanding Climate Change Induced Risk for Investment Pathways in the Bulk Power System

Martha Christino, University of Michigan, Ann Arbor, MI, United States

The increasing frequency and intensity of severe weather events, driven by anthropogenic climate change, poses significant challenges to the reliability of energy systems. Existing research has focused on quantifying the threat to the electricity grid and exploring system operations and capacity expansion needs under multiple future climate realities but lacks analysis of the deep uncertainty created by climate change. Our study aims to explore this deep uncertainty to understand how the value of bulk power system investment pathways change across different future climate realities. We propose a transdisciplinary approach combining dynamically downscaled climate data, an open-source resource adequacy model, and a cross-model analysis framework. We introduce a novel methodology that uses a “risk score” to quantify the robustness of an investment pathway’s resource adequacy contribution under multiple potential future climate realities. We find certain pathways in wind, solar, and storage investments can consistently improve system resource adequacy across future climate realizations, making them robust to the deep uncertainty of future climate states. By developing a risk versus cost frontier, we identify the trade-offs between investment strategies, allowing for informed decision-making amid the complexities introduced by climate change.

4 - Evaluating the Impact of Climate Change and Its Influence on Water Systems on Electricity System Planning in the Western Interconnect

Julia Szinai, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

In the Western US, rising temperatures and changing precipitation patterns from climate change impact the electricity system directly (such as through cooling loads), as well as indirectly through the interdependent water sector (such as through hydropower and energy demand related to water). However, electricity planning models typically ignore these climate and water interactions when considering future investment. We therefore couple detailed water resources and grid capacity expansion models to optimize how the Western Interconnection can both adapt to climate change and develop carbon-free generation by 2050, while accounting for uncertainties from the interactions and climate vulnerabilities of the water sector. We first quantify the range of plausible hydropower and energy demand changes under a large ensemble of future climate scenarios, and then compare the generation and transmission buildout, dispatch, and cost of each climate scenario against a baseline wherein the Western Interconnection reaches carbon-free generation but has a stationary climate. Our results show that cooling and water-related loads increase in all climate scenarios by 2050, which is compounded in most cases by decreases in hydropower generation. The added uncertainty from having to adapt to the worst case climate scenario could increase infrastructure investment costs by up to 7%, highlighting the substantial consequences for grid planning from climate and water interactions and uncertainties.

5 - A National Screening Assessment of Climate Risks to U.S. Hydropower Functions

Jacob Mardian, EPRI, Toronto, ON, Canada, Mark Christian

Hydropower is a unique form of power generation, providing renewable and dispatchable electricity and a range of other services critical to public safety, the economy, and the environment. As climate change alters precipitation patterns and water availability, it is important to consider the impacts on the diverse range of hydropower functions, including power generation, flood control, navigation, recreation, irrigation, municipal water supply, and environmental support. Effectively balancing these multiple, sometimes competing, priorities require effective coordination of available water resources and is dependent on the design of the hydropower facility. Hydropower plants were designed to effectively provide services under historical climate conditions. However, climate change continues to drive increasingly dynamic and non-stationary river conditions, so it is important to review the adequacy of current design and integrate future hydrological regimes into plant modernization. In this study, we develop a series of Hydrological Key Performance Indicators (HKPIs) to project future changes in the ability of ~1,000 hydropower plants in the U.S. to continue providing these critical functions over the near-term (2020 to 2059) and the long-term (2060 to 2099). An HKPI is developed for each function that quantifies projected changes in a relevant hydrological quantity against the plant characteristics describing its ability to cope with those changes. The results show significant variations across hydropower functions, regions, as well as models and scenarios, demonstrating the need for site-specific vulnerability assessments and robust decision-making frameworks under deep uncertainty.

TA33

Building B Level 2 B207

Interpretable LLM/Deep Learning

Invited Session

Health Applications Society

Chair: Shixing Yu, Cornell University, New York, 14850, United States

Co-Chair: Kyra Gan, Cornell Tech, Cornell University, New York, NY, United States

1 - Tbd

Shixing Yu, Cornell/ Cornell Tech, Ithaca, NY, United States

Interpreting the behavior of large language models in critical domains like medical QA requires understanding how internal representations influence final outputs. We propose a framework that extends influence functions, a classical technique from statistics and model interpretability, to analyze how intermediate transformer representations and their compressed sparse autoencoder encodings affect an LLM’s outputs. By inserting a sparse, interpretable bottleneck representation $\$H\$$ (via an encoder $\$E\$$) between an intermediate layer representation $\$R_i\$$ and the model’s final output $\$O_i\$$, we can trace how changes in these hidden states propagate to outputs. This framework combines theoretical tools (Taylor expansions, gradients/Hessians) with practical strategies (autoencoder design for interpretability) to yield a comprehensive approach suitable for research on interpretability and causal analysis of LLMs.

We first recap classical influence functions and draw analogies to intermediate representations. We then develop an influence quantification between an intermediate representation $\$R_i\$$ and model output $\$O_i\$$ using gradient and Hessian-based analyses. Next, we extend this to the two-step path $\$R_i \rightarrow H \rightarrow O_i\$$, leveraging the deterministic SAE mapping $\$H = E(R_i)\$$. We discuss how to design the sparse hidden state $\$H\$$ for maximum interpretability (e.g. disentangled or semantically grounded features). Throughout, we integrate insights from recent work on topic-based influence functions using variational autoencoders

2 - Toward Universal Steering and Monitoring of AI Models

Adityanarayanan Radhakrishnan, MIT and Broad Institute of MIT and Harvard, Cambridge, MA, United States, Daniel Beaglehole, Enric Boix-Adsera, Mikhail Belkin

Modern AI models have learned much of human knowledge, yet understanding of their internal representation of this knowledge remains elusive. Characterizing the structure and properties of this representation will lead to improvements in model capabilities and development of effective safeguards. Building on recent advances in feature learning, we develop an effective, scalable approach for extracting linear representations (vectors and subspaces) of general concepts in large-scale AI models (language models, vision-language models, and reasoning models). We show these representations enable model steering, through which we expose vulnerabilities, mitigate misaligned behaviors, and improve model capabilities. Additionally, we demonstrate that concept representations are remarkably transferrable across human languages and combinable to enable multi-concept steering. Through quantitative analysis across hundreds of concepts, we find that newer, larger models are more steerable and steering can improve model capabilities beyond standard prompting. We show how concept representations are effective for monitoring misaligned content (hallucinations, toxic content). We demonstrate that predictive models built using concept representations are more accurate for monitoring misaligned content than using models that judge outputs directly. Together, our results illustrate the power of using internal representations to map the knowledge in AI models, advance AI safety, and improve model capabilities.

3 - Multimodal Representation Learning for Clinical Decision Support

Sana Tonekaboni, Broad institute of MIT and Harvard, Cambridge, MA, United States

As healthcare data becomes increasingly multimodal—spanning imaging, waveforms, text, and structured records—there is growing interest in methods that can learn from these diverse sources in a unified way. In this talk, I will discuss advances in multimodal representation learning, focusing on how they can help bridge the gap between heterogeneous clinical inputs and downstream decision-making. I'll outline key challenges in aligning and interpreting information across modalities, and highlight design choices that affect generalization and interpretability in clinical settings. To ground the discussion, I'll present our recent work integrating electrocardiogram (ECG) signals and cardiac MRI to learn joint representations that improve cardiovascular risk assessment. Our findings show how shared and modality-specific patterns can be disentangled to reveal clinically meaningful insights, such as structural-functional relationships in the heart.

4 - Representation Learning for Cell and Tissue Biology: from Multimodal Integration to Function

Xinyi Zhang, Broad Institute of MIT, Boston, MA, United States

Biological processes involve complex interactions across different scales, from intracellular changes in gene expression and chromatin organization to intercellular communication and tissue organization. Comprehensive understanding of cell and tissue states as well as disease mechanisms requires computational methods that can integrate diverse data types. First, we introduce a graph-based autoencoder that integrates gene expression, spatial coordinates, and cellular imaging into a joint representation. Applied to a mouse model of Alzheimer's disease, this approach reveals spatio-temporal disease patterns and associated nuclear and transcriptional changes. To address the lack of scalability in multimodal and sequencing-based assays, we develop unsupervised image autoencoder frameworks that extract cell state information from imaging data and enables the identification of cell state and tissue organization based on imaging. The proportions and spatial organization of these states predict diagnosis, phenotype, and mutation status, providing scalable biomarkers and insight into chromatin and microenvironmental involvement.

Building on the richness of cellular images, we developed PUPS (Predictions of Unseen Proteins' Subcellular localization), which combines protein language models with image inpainting to predict protein localization. PUPS overcomes experimental limitations in scalability and reveals associations between protein localization variability and biological function. Finally, we introduce APOLLO, which disentangles shared and modality-specific information in multimodal datasets. Applied to transcriptomic and imaging data, APOLLO identifies shared gene pathways and links localization variability to cellular compartment morphology. Together, these frameworks enable integration of diverse modalities, advancing our understanding of cell states, tissue organization, and disease.

TA34

Building B Level 2 B208

Learning with Humans, AI, and Limited Data

Invited Session

OPT: Machine Learning

Chair: Madeleine Pollack, Massachusetts Institute of Technology, Cambridge, MA, 02138, United States

Co-Chair: Swati Gupta, Sloan School of Management, MIT, Cambridge, MA, United States

Co-Chair: Jai Moondra, Georgia Institute of Technology, Atlanta, GA, 30332, United States

1 - Boosting Kidney Transplant Access: Interpretable AI to Discover Opportunities for Xenografted and Cryopreserved Organs

Thomas Dailak, Massachusetts Institute of Technology, Cambridge, MA, United States, Ju Young Lee, Madeleine Pollack, Prateek Gautam, David Cron, Mehmet Toner, Korkut Uygun, Heidi Yeh, Swati Gupta

Recent breakthroughs in cryopreservation and xenotransplantation offer new possibilities for expanding the supply of viable donor kidneys. However, it remains unclear where in the current transplantation pipeline these innovations would have the maximum impact. In this work, we use interpretable AI methods to identify systemic inefficiencies and opportunities for placing either cryopreserved or xenografted organs. We focus on two key questions: (i) can we quantify the supply-demand gap across different patient groups and regions, using organ quality as a proxy? and (ii) can we develop predictive tools to identify regions where a given organ is most likely to be accepted, based on local clinical behavior? We will present our analysis of over 8 million donor-patient match-run offers from UNOS (2016–2019), using multimodal predictive models for three demographically diverse regions. We introduce a new metric for quantifying the supply demand gap in this complex pipeline, and show organ profiles that would be most useful targets for xenografting. We further identify areas of model uncertainty based on organ quality, data imbalance and the region of performance, and discuss the multi-criteria considerations for organ placement using cryopreservation. Our interpretable methods include an extension of SHAP analysis for multimodal models, which further reveals key donor, recipient, and text-feature interactions that are predictive of offer acceptance. This presentation is based on joint work with Ju Young Lee, Madeleine Pollack, Prateek Gautam, David Cron, Mehmet Toner, Korkut Uygun, Heidi Yeh, and Swati Gupta.

2 - AI Meets Peer Review: The Good, the Bad, and the Ugly

Nihar Shah, Carnegie Mellon University, Pittsburg, PA, United States

As AI becomes increasingly integrated into academic workflows, their influence on scientific peer review is both promising and concerning. This talk will explore three facets of this evolving intersection.

The Good: AI's ability to execute aspects of peer review that are challenging for human reviewers.

The Bad: Vulnerabilities in the review process to fraud such as collusion rings.

The Ugly: Reviewers using AI to generate reviews, and detection of such usage.

3 - Q-Learning with Posterior Sampling

Priyank Agrawal, Columbia University, New York, NY, United States, Shipra Agrawal

Bayesian posterior sampling techniques have demonstrated superior empirical performance in many exploration-exploitation settings. However, their theoretical analysis remains a challenge, especially in complex settings like reinforcement learning. In this talk, I will introduce Q-Learning with Posterior Sampling (PSQL), a simple Q-learning-based algorithm that uses Gaussian posteriors on Q-values for exploration, akin to the popular Thompson Sampling algorithm in the multi-armed bandit setting. I will show that in the tabular episodic MDP setting, PSQL achieves a regret bound of $\tilde{O}(\sqrt{SAT})$, closely matching the known lower bound of $\Omega(\sqrt{SAT})$. Here, S, A denote the number of states and actions in the underlying Markov Decision Process (MDP), and $T=KH$ with K being the number

of episodes and H being the planning horizon. This work provides several new technical insights into the core challenges in combining posterior sampling with dynamic programming and TD-learning-based RL algorithms, along with novel ideas for resolving those difficulties. I hope this will form a starting point for analyzing this efficient and important algorithmic technique in even more complex RL settings.

4 - Cost-Averse Learning with Mixed-Fidelity Bandit Feedback

Madeleine Pollack, Massachusetts Institute of Technology, Cambridge, MA, United States

In today's data-driven world, decision-makers face a complex landscape of diverse data sources. They must navigate a mosaic of high- and low-fidelity information, each with its own cost implications and impact on decision reliability. The quality of information directly affects operational performance, so care must be taken to de-bias, de-noise, and properly integrate data from these differing-fidelity sources. This challenge is particularly acute when faulty information can lead to significant adverse outcomes in systems such as human-AI pipelines in decision-making (e.g., content moderation, labeling, fact-checking), prescriptive medicine, supply chains, and other applications. We introduce a new bandit model to study the trade-off between quality of feedback and the cost of the feedback mechanism. We show that without distinguishing between the feedback mechanisms generating the observed rewards, algorithms like UCB can incur linear regret. We propose a new Combined-Learning UCB algorithm that can calibrate the initially unknown fidelities of different feedback mechanisms. We show that in certain settings, this algorithm outperforms a mechanism-conscious version of UCB1. We finally showcase the impact of our algorithms on a dataset from a call center, where experienced callers can convert a higher call volume into sales, compared to new callers by enabling cost-aware decision-making in dynamic environments. This approach contributes to more reliable and robust decision-making, particularly in settings where data quality varies widely.

5 - Practical Contextual Bandits for Large-Scale Structured Discrete Constrained Optimization Problems

Pavithra Harsha, IBM Research, Yorktown Heights, NY, United States, Chitra Subramanian, Naoki Abe, Shivaram Subramanian, Amadou Ba, Kevin Fernandez, Mauricio Longinos

We address the contextual bandit problem for high-dimensional combinatorial action spaces involving a class of structured discrete constrained optimization problems. In this setting, key quantities within the optimization formulation, such as performance indicators contributing to the objective or constraints, must be estimated from data during the trial sequence of optimized actions. These problems frequently arise in the many operations management domains including IT resource allocation and retail assortment price optimization. We propose a novel, practical, and transparent approach based on general-purpose regression oracles, leveraging Inverse Gap Weighting (IGW) for seamless integration within an optimization framework. IGW sampling is efficiently managed by: (a) a column generation reformulation of the underlying Mixed Integer Programming (MIP) model which allows for flexible lower-level predictors, causal coherence, and efficient representation of large action spaces; (b) a diverse solution pool generation to balance the exploration-exploitation trade-off in large action spaces. To address non-smoothness in the reward function due to optimization constraints, we incorporate a risk-averse phased learning strategy. We validate our approach on a real world auto-scaling problem in IT automation, achieving a significant reduction in cumulative regret through skillful exploration, with additional gains from risk-averse methods that effectively manage constraint violations.

TA35

Building B Level 2 B209

Cooperation and Partnership in Public Services

Invited Session

Public Sector OR

Chair: Priscila de Azevedo Drummond, Northeastern University, Boston, United States

1 - The Value of Incorporating Uncertainty in School District Design

Aysu Ozel, Kennesaw State University, Kennesaw, GA, United States, Karen Smilowitz

Long-term public school district design decisions, such as opening or closing schools, drawing attendance boundaries, and placing programs in schools, can significantly shape students' educational experiences. While such decisions aim to provide students with continuity, they are often made under uncertainty in student enrollment. We explore the value of incorporating uncertainty into school district design through a family of models that differ in their timing of long-term decisions and how they account for uncertain enrollment. We develop scenario-based models that support multi-year school district design under enrollment uncertainty, with decision structures ranging from fixed, one-time decisions to adaptive, scenario-dependent ones. To evaluate these models, we generate realistic enrollment scenarios using historical enrollment data and census-derived spatial distributions for representative simulated school districts. These scenarios reflect structured patterns in student progression and capture both aggregate trends and geographic variation in enrollment changes. Our analysis examines the trade-offs between model complexity and solution robustness. In particular, we assess when and how stochastic and dynamic models may provide meaningful advantages over simpler approaches, given the structured yet uncertain nature of enrollment shifts. This work contributes to the broader literature on decision-making under uncertainty by offering insights into the conditions under which uncertainty-aware models can enhance the effectiveness of school district design decisions.

2 - Perspectives on OR for Societal Good, a Review of the Last 10 Years, and a Call for a Research Agenda for the Next 10 Years

Matt Kammer-Kerwick, The University of Texas at Austin, Austin, TX, United States

This presentation will provide a high-level overview of 10 years of collaborative research in Operations Research (OR) for the Social Good and a call for a new or reaffirmed research agenda for the next decade.

As context, in 2017 National Science Foundation funded a workshop on promising research directions for applications of operations research and data analytics toward the disruption of illicit supply networks like human trafficking as a specific social good thematic focus. The workshop was funded by the NSF's Operations Engineering (ENG) and the Law & Social Sciences Program (SBE) under grant # CMMI-1726895. Participants addressed the opportunity to apply advances from the fields of operations research, management science, analytics, machine learning, and data science toward the development of disruptive interventions against illicit networks.

This presentation will briefly review the proposed research agenda from that workshop and some of the research streams that emerged. Discussion will be invited about an agenda for the next 10 years.

3 - Collaboration in Antitrafficking Efforts: a Network Interdiction Problem

Priscila de Azevedo Drummond, Northeastern University, Boston, MA, United States, Kayse Maass

Labor trafficking is a ubiquitous issue in the U.S. affecting various sectors from agricultural workers to the assessment of imported goods. combating labor trafficking has many different facets. The efforts to combat and prevent this crime involve numerous agencies, each with different areas of expertise and communities they serve. These agencies often have varying values and operational procedures, which can complicate collaboration. While teamwork has been shown to enhance the effectiveness of labor trafficking mitigation strategies, data supporting this effectiveness is scarce, and thorough evaluations are even rarer. Moreover, collaboration requires time and resources, and most anti-human trafficking operations run under strict budget constraints. The lack of data hinders our ability to develop responses based on historical evaluations. In these situations, analytical tools like Network Indicator Models can provide insights to help us understand how collaboration functions within this context.

4 - Developing a Systems-Focused Tool for Modeling Lung Cancer Screening Resource Needs in the Veteran's Affairs Health Service

Kira Woodhouse, Northeastern University, Boston, MA, United States

Within past decades, the medical field has experienced a rapid expansion in the number of available cancer screening interventions. While it is known that cancer screenings have many benefits on both a patient- and hospital-level, the question of how to implement them in these settings remains. In a partnership between the University of Michigan and the Veteran's Affairs (VA) Lung Precision Oncology Program, we examined the existing structures of Lung Cancer Screening (LCS) in the VA system through informational interviews and data collection. This was followed by collaborative system-mapping exercises with providers and the development of

calculation tools, both deterministic and stochastic. Such tools consider the patient volume, provider volume, provider tasks, and the cyclical nature of screening to produce estimations of the required staff hours for a scenario. Finally, we released these calculation tools to current staff of the LCS programs nationwide, aiding them in making informed decisions on both staffing levels and staff responsibilities.

5 - The Impact of Access to Food Retail Stores on Household Food Waste

Vijaya Chebolu-Subramanian, Drexel University, Philadelphia, PA, United States, Jennifer Quinlan

Food waste in the United States is estimated to be between 30–40% of the total food supply, with households responsible for the majority of waste along the downstream segment of the supply chain. This waste is valued at approximately \$240 billion annually, representing both economic and environmental loss. The objective of this study was to examine the impact of access to food retail on household food waste across income levels in the United States. According to the U.S. Department of Agriculture, two primary indicators of household retail access are distance to food stores and vehicle ownership. Prior research suggests that access to retail stores can influence shopping frequency, impulse purchases, and overbuying behavior—all of which are linked to food waste. We administered two nationwide surveys ($n=1,103$; $n=816$) during 2023–2024, capturing household food waste patterns and shopping behaviors. Using ordered probit regression models, we found that greater distance to stores was significantly associated with increased food waste among lower-income households. Additionally, lack of vehicle ownership correlated with higher food waste across all income groups. Importantly, for lower-income households, the combination of long distances and no car ownership had a compounding effect on food waste. This study contributes to a deeper understanding of access barriers to retail stores—specifically distance and transportation—as critical factors influencing household food waste, particularly among lower-income populations. The findings highlight the need for further research and targeted interventions to reduce these barriers and support food waste reduction efforts.

TA36

Building B Level 2 B210

Recent Advances in Nonconvex Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Salar Fattahi, University of Michigan, Ann Arbor, MI, United States

1 - Active Set Identification and Rapid Convergence for Primal-Dual Degenerate Problems

Mateo Diaz, Johns Hopkins University, Baltimore, MD, United States, Pedro Izquierdo Lehmann, Haihao Lu, Jinwen Yang

Primal-dual methods for solving optimization problems with functional constraints often exhibit a distinct two-stage behavior. Initially, they converge towards a solution at a sublinear rate. Then, after a certain point, all iterates accurately identify the set of active constraints, after which their convergence accelerates to a much faster linear rate. Theory characterizing this phenomenon spans more than three decades. However, most existing work only guarantees eventual active set identification and relies heavily on nondegeneracy conditions, such as strict complementarity, which frequently fail to hold in practice. We introduce mild conditions on the problem geometry and algorithm under which this phenomenon provably occurs. Our guarantees are entirely nonasymptotic and, importantly, do not require any nondegeneracy assumptions. We demonstrate that our framework encompasses several widely-used algorithms, including the proximal point method, primal-dual hybrid gradient method, alternating direction method of multipliers, extragradient method, and optimistic gradient method.

2 - A Hybrid Newton-CG-MR Method with Worst Case Complexity Guarantees for Nonconvex Optimization

Michael O'Neill, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Chengze Xie

Inexact Newton methods have a long history of popularity in large scale, nonlinear optimization. We propose a new Newton type algorithm which uses a combination of the conjugate gradient (CG) and minimum residual (MR) solvers for linear systems to generate search directions. The new method outperforms both Newton-CG and Newton-MR methods on a wide range of test problems while maintaining optimal worst-case complexity guarantees for nonconvex optimization.

3 - Solving Convex Quadratic Optimization with Indicators over Graphs with Small Treewidth

Aaresh Bhathena, University of Michigan, Ann Arbor, MI, United States, Salar Fattahi, Andres Gomez, Simge Kucukyavuz

This paper studies convex quadratic optimization problems where each continuous variable is coupled with an indicator variable. Our focus is on the structured case in which the matrix Q defining the quadratic term is positive definite and whose sparsity pattern corresponds to the adjacency matrix of a graph with small treewidth. We introduce a dynamic programming algorithm that solves this problem in linear time with respect to n , under the assumptions of a fixed condition number of Q , fixed treewidth, and linear volume growth of the underlying graph. Computational experiments on synthetic data demonstrate that our algorithm significantly outperforms state-of-the-art mixed-integer quadratic programming solvers on instances with over 20,000 indicator variables. To demonstrate the practical utility of the method, we develop a novel framework for joint forecasting and anomaly detection by extending exponential smoothing to handle time series data with outliers. This framework is applied to various real-world instances from the Numenta Anomaly Benchmark, which includes traffic and AWS CloudWatch datasets.

4 - Proximal Random Reshuffling Under Local Lipschitz Continuity

Xiaopeng Li, Columbia University, New York, NY, United States, Cedric Jozs, Lexiao Lai

We study proximal random reshuffling (PRR) for minimizing the sum of locally Lipschitz or locally smooth functions and a proper lower semicontinuous convex function without assuming coercivity or the existence of limit points. The algorithmic guarantees pertaining to near approximate stationarity rely on a new tracking lemma linking the iterates to trajectories of conservative fields. One of the novelties in the analysis consists in handling conservative fields with unbounded values. In the locally smooth case, it improves the known convergence rate from nearly $\mathcal{O}(k^{-1/4})$ to nearly $\mathcal{O}(k^{-1/2})$.

5 - Parameter-Free Second-Order Methods for Min-Max Optimization

Ali Kavis, University of Texas at Austin, Austin, TX, United States

I will talk about an adaptive, line-search-free second-order methods with optimal rate of convergence for solving convex-concave min-max problems. By means of an adaptive step size, our algorithms feature a simple update rule that requires solving only one linear system per iteration, eliminating the need for line-search or backtracking mechanisms. Specifically, we base our algorithms on the optimistic method and appropriately combine it with second-order information. Moreover, distinct from common adaptive schemes, we define the step size recursively as a function of the gradient norm and the prediction error in the optimistic update. We first analyze a variant where the step size requires knowledge of the Lipschitz constant of the Hessian. Under the additional assumption of Lipschitz continuous gradients, we further design a parameter-free version by tracking the Hessian Lipschitz constant locally and ensuring the iterates remain bounded. We also evaluate the practical performance of our algorithm by comparing it to existing second-order algorithms for minimax optimization. Inspired by the adaptive design of the step size, we propose a heuristic initialization rule that performs competitively across different problems and scenarios and eliminates the need to fine tune the step size.

TA37

Building B Level 2 B211

Decision-Focused Learning: Methodology and Applications

Invited Session

OPT: Optimization Under Uncertainty

Chair: Tito Homem-de-Mello, Universidad Adolfo Ibañez, Santiago, Chile

1 - Neural Decision Rule for Constrained Contextual Stochastic Optimization

Rui Gao, The University of Texas at Austin, Austin, TX, United States

Contextual stochastic optimization is a powerful paradigm for data-driven decision-making under uncertainty, where decisions are tailored to contextual information revealed prior to decision making. Recent advances leverage neural networks to learn expressive decision rules mapping contexts to decisions. However, enforcing feasibility under complex constraints remains a core challenge, as neural architectures do not inherently satisfy constraints in general. Existing approaches often incur significant computational overhead and lack theoretical guarantees.

In this paper, we propose a principled framework for training neural decision rules under general constraints via a single-loop algorithm that solves the augmented Lagrangian minimax problem. Our method eliminates the need for iterative projection layers or nested optimization loops, and provides provable guarantees on generalization, constraint violation, and suboptimality. Empirical results on constrained contextual decision tasks demonstrate that our approach outperforms state-of-the-art baselines in both efficiency and solution quality.

2 - Pessimistic Bilevel Optimization Approach for Decision-Focused Learning

Bernardo Pagnoncelli, SKEMA Business School, Lille, France, Diego Jimenez, Hande Yaman

The recent interest in contextual optimization problems, where randomness is associated with side information, has led to two primary strategies for formulation and solution. The first, estimate-then-optimize, separates the estimation of the problem's parameters from the optimization process. The second, decision-focused optimization, integrates the optimization problem's structure directly into the prediction procedure. In this work, we propose a pessimistic bilevel approach for solving general decision-focused formulations of combinatorial optimization problems. Our method solves an approximation of the pessimistic bilevel problem using a specialized cut generation algorithm. We benchmark its performance on the 0-1 knapsack problem against estimate-then-optimize and decision-focused methods, including the popular SPO+ approach. Computational experiments highlight the proposed method's advantages, particularly in reducing out-of-sample regret.

3 - Predict and Optimize in Energy Markets

Pascal Van Hentenryck, ISyE Georgia Tech, Atlanta, GA, United States

This presentation reviews the fundamentals of Predict and Optimize for Energy markets applications. It discusses their benefits on practical applications, computational issues that arise in practice, and how they can be addressed. In particular, it reviews how primal and dual optimization proxies can be helpful in speeding up decision-focused learning with negligible loss in accuracy.

4 - A Fully First-Order Method for Decision-Focused Learning

Kai Wang, Georgia Institute of Technology, Atlanta, GA, United States

Decision-focused learning (DFL) aims to train predictive models by optimizing downstream decision quality rather than traditional predictive accuracy. While recent algorithms have integrated optimization directly into the training pipeline to achieve decision-focused learning, they often rely on second-order derivatives or surrogate losses, limiting their scalability and applicability in large-scale or constrained settings. In this talk, I will introduce a fully first-order method for DFL, grounded in recent advances in bilevel optimization algorithms. Our approach enables efficient end-to-end training using only first-order gradients by reformulating the decision problem with a smooth penalty, allowing gradients to backpropagate through the optimization layer without requiring differentiation through exact solvers. I will discuss both the theoretical convergence guarantees of our method and the empirical results demonstrating strong computation advantage and performance on large-scale problems. This work on fully first-order DFL advances the scalability and practicality of decision-focused learning by eliminating key computational bottlenecks.

5 - An Adaptive-Partition Method for Decision-Focused Learning

Tito Homem-de-Mello, Universidad Adolfo Ibáñez, Santiago, Chile

Data-driven optimization problems with contextual information have given rise to a number of new approaches that integrate the estimation and optimization steps. In this talk we consider one such approach, which formulates the problem of estimating the parameters for the learning method as a bilevel optimization model with multiple lower levels. While such a formulation is provably convergent to the optimal parameters, the model is computationally challenging. We discuss an approach based on the Adaptive Partition Method (APM) introduced in the literature for two-stage stochastic optimization. With the APM approach, the number of lower level problems can be reduced considerably. We illustrate the method with an application to the optimization of microgrids in power systems.

TA38

Building B Level 2 B212

Recent Progress in Optimization Software I

Invited Session

OPT: Computational Optimization and Software

Chair: Hans Mittelman, Arizona State University, Tempe, AZ, United States

1 - Accelerating Low-Rank Factorization-Based Semidefinite Programming Algorithms on GPU

Qiushi Han, UNIVERSITY OF ILLINOIS AT URBANA - CHAMPAIGN, Urbana, IL, United States, Zhenwei Lin, Hanwen Liu, Caihua Chen, Qi Deng, Dongdong Ge, Yinyu Ye

In this paper, we address a long-standing challenge: how to achieve both efficiency and scalability in solving semidefinite programming problems. We propose acceleration techniques for a wide range of low-rank factorization-based first-order methods using GPUs, making the computation much more efficient and scalable. To illustrate the idea and effectiveness of our approach, we use the low-rank factorization-based SDP solver, LoRADS, as an example, which involves both the classic Burer-Monteiro method and a novel splitting scheme with a starting logarithmic rank. Our numerical results demonstrate that the accelerated GPU version of LoRADS, cuLoRADS, can solve huge-scale semidefinite programming problems with remarkable efficiency. By effectively leveraging GPU computational power, cuLoRADS exhibits outstanding performance. Specifically, it can solve a set of MaxCut problems with $10^7 \times 10^7$ matrix variables in 10 seconds to 1 minute each on an NVIDIA H100 GPU with 80GB memory, whereas previous solvers demonstrated the capability of handling problems of this scale, required at least dozens of hours per problem on CPUs. Additionally, cuLoRADS shows exceptional scalability by solving 1) a MaxCut problem with a 170 million $\times$ 170 million matrix variable and 2) a Matrix Completion problem with a 20 million $\times$ 20 million matrix variable and approximately 200 million constraints, both in a matter of minutes.

2 - Gpu-Based Mathematical Programming

Haihao Lu, MIT, Cambridge, MA, United States, Jinwen Yang

In this talk, I will provide an overview and recent updates on my group's research in GPU-based mathematical programming, with a focus on linear programming, convex quadratic programming, semidefinite programming, and potentially other related topics.

3 - Advances in Nvidia CuOpt : Open Source Optimization on GPUs

Alex Fender, NVIDIA, Santa Clara, CA, United States

cuOpt is NVIDIA's solver for decision optimization, built for accelerated architectures. This talk introduces its core capabilities and highlights recent advancements in both design and integration.

A major milestone this year is the first open-source release of cuOpt under the Apache 2.0 license, supported by an active ecosystem of researchers, solvers and modeling tools. We'll explore how cuOpt leverages recent hardware to push the boundaries of decision optimization performance. Attendees will gain a deeper understanding of how cuOpt bridges the gap between research on parallel algorithms and real-world systems.

4 - Distributed Hybrid Compute for Large-Scale MIP and LP Optimization

Mark Ostroot, Simpleroose, St. Louis, MO, United States

Recent advances in GPU computing have opened new possibilities for accelerating mixed-integer and linear programming. At SimpleRose, we've been exploring distributed CPU/GPU hybrid compute architectures that leverage NVIDIA's cloud-scale hardware to push the boundaries of what's solvable. We'll share early insights from our ongoing work, including integration efforts inspired by frameworks like cuOpt, and discuss the challenges and opportunities in scaling optimization solvers across heterogeneous compute environments.

5 - Latest Optimization Benchmark Results

Hans Mittelman, Arizona State University, Tempe, AZ, United States

A selection of our optimization software benchmarks will be presented.

TA39

Building B Level 2 B213

Progress in Mixed-integer Nonlinear Programming

Invited Session

OPT: Integer and Discrete Optimization

Chair: Aleksandr Kazachkov, University of Florida, Gainesville, United States

1 - Cutting Planes for Mixed-Integer Semidefinite Programming

Jan Kronqvist, KTH Royal Institute of Technology, Stockholm, Sweden, Johanna Skåntorp

Here we focus on Mixed-Integer Semidefinite Programming (MISDP), and specifically solving MISDP by outer approximation techniques. We present an alternative interpretation of some known types of cutting planes, and some new techniques to derive cuts with more desirable properties. We also present a new primal heuristic that also builds on an outer approximation concept that can both find feasible solutions and produce cutting planes for the main algorithm.

2 - Revisiting Composite Relaxations from a Lifting Lens

Haisheng Zhu, Purdue University, West Lafayette, IN, United States, Taotao He, Mohit Tawarmalani

We develop a new relaxation for nonlinear expressions in factorable programming. This relaxation generalizes composite relaxation by using a better abstraction of the inner function, which is constructed by the perturbation of the inner function. In limiting case, our relaxation converges to the convex hull of the graph of $f_1(x_1) f_2(x_2)$ over $X_1 \times X_2$ when f_1 and f_2 are convex. Our relaxation can improve factorable programming by introducing few auxiliary variables.

3 - Minlp for Regularized Symbolic Regression

Dimitrios Fardis, Georgia Institute of Technology, Atlanta, GA, United States, Radhakrishna Tumbalam Gooty, Nikolaos Sahinidis

Over the past decade, symbolic regression has been approached through MINLP formulations. We investigate various regularization metrics in the context of symbolic regression. We present extensive computational experimentation with the different metrics as we evaluate them for solvability of the resulting MINLP and for their ability to produce symbolic regression models that are simple and accurate.

4 - Globally Solving Assortment Optimization Problems via Multilinear Optimization Methods

Yating Zhang, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Taotao He, Zhongqi Wu, Huan Zheng

In assortment planning, the goal is to select a subset of products that maximizes expected revenue, assuming that consumers' stochastic choice behavior has already been estimated. This leads to a deterministic nonlinear 0-1 programming problem. In this talk, we investigate two such problems through the lens of multilinear optimization. For the correlated choice model, we use convexification techniques from fractional and multilinear optimization to obtain bounds-free MIP formulations, which are tighter than those obtained using Big-M approaches. For the ranking-based choice model, we reformulate the problem as a 0-1 multilinear optimization problem, and leverage consumers' rankings to derive multiple classes of valid inequalities. We also provide numerical experiments to demonstrate the efficiency of our formulations.

5 - On Local Search in Bilevel Mixed-Integer Programming

Eneko Clemente, University of Zurich, Zurich, Switzerland, Oleg Prokopyev

We consider Bilevel Mixed-Integer Linear Programs (Bilevel MILPs) in which a leader makes a decision, anticipating that a follower subsequently solves an optimization problem that depends on the leader's decision. Local search in Bilevel MILPs encounters two main challenges: (i) similar to many combinatorial optimization problems, local search may require an exponential number of improving steps to converge in the worst case, and (ii) each evaluation of the leader's objective function value requires solving the follower's problem, which can be inherently difficult. We address these two challenges by exploring generalized notions of local optimality that leverage the structural properties of BMILPs.

6 - Simple Randomized Rounding for Max-Min Eigenvalue Augmentation

Haeseong Yang, University of Pittsburgh, Pittsburgh, PA, United States

We consider the max-min eigenvalue augmentation problem: given $n \times n$ symmetric positive semidefinite matrices $M, A_1, \dots, A_m$ and a positive integer $k < m$, the goal is to choose a subset $I \subset \{1, \dots, m\}$ of cardinality at most k that maximizes the minimum eigenvalue of the matrix $M + \sum_{i \in I} A_i$. The problem captures both the Bayesian E-optimal design and maximum algebraic connectivity augmentation problems. In contrast to the existing work, we do not assume that the augmentation matrices are rank-one matrices, and we focus on the setting in which $k < n$. We show that a simple randomized rounding method provides a constant-factor approximation if the optimal increase is sufficiently large, specifically, if $\text{OPT} - \lambda_{\min}(M) = \Omega(R \ln k)$, where OPT is the optimal value, and R is the maximum trace of an augmentation matrix. To establish the guarantee, we derive a matrix concentration inequality that is of independent interest. The inequality can be interpreted as an intrinsic dimension analog of the matrix Chernoff inequality for the minimum eigenvalue of a sum of independent random positive semidefinite matrices; such an inequality has already been established for the maximum eigenvalue, but not for the minimum eigenvalue.

TA40

Building B Level 2 B214

Customer Behavior and Demand Modeling

Contributed Session

Contributed

Chair: Fatima Ulubayova, McMaster University, 1261 Main Street West, Hamilton, L8S 1C4, Canada

1 - Using Sensor Technology to Evaluate the Effects of a Price Promotion on Store Arrival Times

Ignacio Inostroza-Quezada, Universidad de los Andes, Santiago, Chile, Leonardo Epstein, Ronald Goodstein

Store managers frequently conduct marketing actions (MAs) to boost store traffic. Thus, measuring this traffic is crucial for assessing the effectiveness of such initiatives. Recently, store managers have gained access to sensors that record data continuously, enabling a more in-depth analysis of how their actions affect consumer behavior. In this talk, we present a novel approach that uses shopper-specific arrival times from periods outside a MA to build a counterfactual baseline for store traffic during this MA.

This baseline (i) assumes that the MA has not taken place, and (ii) is essential because a store cannot operate simultaneously with and without the MA. Next, we compare observed traffic data with the counterfactual baseline during the MA to evaluate its effect on store traffic. Finally, we assess the uncertainty of this effect during the MA. We illustrate our approach with data from an actual two-week price promotion, which showed an increase of 2880 store arrivals attributed to the promotion. Additionally, we complement our results with visual plots that help managers (i) examine the evolving effects of the promotion and (ii) understand the influence of time-varying covariates on consumer behavior.

2 - Optimal Job Recommendation Policies in Online Platforms with No-Shows and Stochastic Arrivals

HEDIEH SAZVAR, Northwestern University, Evanston, IL, United States, Seyed Iravani, Sanjay Mehrotra

The increasing reliance on online platforms to match job opportunities with agents highlights the need for efficient job recommendation policies. This paper aims to optimize job recommendations to maximize platform revenue. Our model considers different preferences and commitment probabilities for each agent-job pair, as well as varying number of agents required for different jobs. We propose a job recommendation model for implementation on a matching platform, where agents arrive sequentially to the platform (i.e., log in one after another). The objective is to maximize the revenue earned when agents complete their jobs.

We develop a heuristic algorithm to generate near-optimal recommendation policies under stochastic agent arrivals, and introduce an efficient objective function evaluation method to support large-scale implementation. Moreover, for platforms that can control the arrival sequence of agents, we identify the optimal sequence and derive the corresponding optimal recommendation policy to further improve platform revenue.

3 - A Rolling Recruitment Process Under Applicant Stochastic Departures

Chenyin Gong, Hong Kong University of Science and Technology, Hong Kong, China, People's Republic of, Qing Li

We study a rolling recruitment process in which applicants leave the system stochastically. Applicants arrive randomly over time, and each applicant is available for a random amount of time after they arrive. In each period, the recruiter must decide whether to stop or to wait. If they stop, they need to determine how many offers to make and whom to make offers to, and the applicants who do not receive an offer will leave the system. If the recruiter waits, more applicants will be available in the next period, while some who arrived earlier will leave. We model the process as a large-scale optimal stopping problem and show how the applicant qualifications, measured by scores, affect the recruiter's optimal policy. We find that the optimal stopping rule for each applicant's score is a two-threshold policy. If the score exceeds the higher threshold, the recruiter stops and makes an offer to the applicant and possibly to others. If the score falls below the lower threshold, the recruiter also stops but makes no offer to the applicant. If the score is in between the two thresholds, the recruiter waits. We further explore the impact on an applicant's likelihood of receiving an offer if their competitors become more qualified. We show that an applicant may be disadvantaged when competing against stronger applicants; however, they may also benefit from such competition. These results can be extended to allow multiple scores to change simultaneously.

4 - Optimal Inventory Disclosure and Pricing when Consumers Experience Stock-Out Inconvenience

Ceren Gultekin, Baruch College, CUNY, Graduate Center, New York, NY, United States, Tolga Aydinliyim, Qiuyu Fan

Retailers employ various non-price tactics to discourage consumers' strategic purchase deferrals and stimulate full price demand ahead of markdowns. One such strategy is to use availability indicators, which may suggest future stock-outs. These scarcity signals are commonplace, but recent studies argue they may backfire as consumers avoid transacting with limited-availability products. While these studies make empirical observations regarding how inventory disclosure affects sales, they fall short of prescribing actionable insights. We fill this gap by modeling optimal inventory disclosure and pricing decisions for retailers facing forward-looking consumers. We find that selective inventory disclosure (share low & mask high stock levels) remains optimal even when consumers experience an inconvenience cost at stock-out events. However, this does not imply that retailers should employ more inventory masking. In contrast, we show that inventory sharing must be more robustly utilized, i.e., at even higher stock levels relative to when consumers do not incur a stock-out inconvenience cost.

5 - Joint Estimation of the Arrival Rate and Customer Taste Coefficients from Censored Transactional Data

Weikun Xu, Washington State University, Pullman, WA, United States, Xinchang Wang

This work aims to jointly estimate the arrival rate of customers to a market and the nested logit model that forecasts hierarchical customer choices from an assortment of products. The estimation is based on censored transactional data, where lost sales are not recorded. The goal is to determine the arrival rate, customer taste coefficients, and nest dissimilarity parameters that maximize the likelihood of the observed data. The problem is formulated as a maximum likelihood estimation model that addresses two prevailing challenges in the existing literature: Estimating demand from *data with unobservable lost sales* and capturing *customer taste heterogeneity arising from hierarchical choices*. However, the model is intractable to solve or analyze due to the nonconcavity of the likelihood function in both taste coefficients and dissimilarity parameters. We characterize conditions under which the model parameters are identifiable. Our results reveal that the parameter identification is influenced by the *diversity of products and nests*. We also develop a sequential minorization-maximization algorithm to solve the problem, by which the problem boils down to solving a series of convex optimization models with simple structures. Then, we show the convergence of the algorithm by leveraging the structural properties of these models. We evaluate the performance of the algorithm by comparing it with benchmarks, using both synthetic and real data. Our findings show that the algorithm consistently outperforms the benchmarks. Moreover, our algorithm is particularly effective in estimating nested logit models with low dissimilarity parameters, yielding higher profitability compared to the benchmarks.

6 - Temporal Decision Making in High-Demand Sales Events with Limited Inventory

Fatima Ulubayova, McMaster University, Hamilton, ON, Canada, Yun Zhou, Mahmut Parlar

In this paper, we explore a critical challenge in high-demand sales events for limited products, where the imbalance between product availability and consumer demand introduces complex strategic behavior. We introduce a general model for such sales events, where a finite number of product units q are allocated among n arriving customers, with $q < n$. Our analysis considers scenarios involving one, two, and n competing customers, under both identical and heterogeneous arrival time distributions. As the number of competitors increases, the complexity of the allocation dynamics escalates. Using order statistics, we frame the problem as a benefit maximization problem from the customer's perspective and characterize the optimal arrival time based on an optimality condition. We further analyze the problem through a game-theoretic setting. For the case of two strategic customers, we model the scenario as a static game with complete and incomplete information with discrete and continuous strategies. In this setting, each player's effort affects their random arrival times. These efforts reflect the cost and inconvenience of attempting to arrive early, which may include opportunity costs, travel expenses, and the inherent hassle of the process. We prove the existence and uniqueness of the Nash equilibrium under general conditions and derive the explicit equilibrium in the special case where effort costs are zero for all players. Our results provide insight into the strategic behavior of consumers in high-demand sales settings and highlight the role of timing and effort in securing limited products.

TA41

Building B Level 3 B301

Data-Driven Healthcare Operations

Invited Session

Health Applications Society

Chair: Wilson Lin, Santa Clara University, Santa Clara, CA, United States

1 - Cost-Effectiveness of a Lethal Ovitrap Network for Dengue Fever Prevention and Control

Yvonne Huijun Zhu, National University of Singapore, Singapore, Singapore, Joel Aik, Shuzhen Sim, Joel Goh

We conduct a system-level analysis of the cost-effectiveness of lethal ovitraps for dengue fever prevention, assessing both epidemiological outcomes and operational costs. The health benefits are captured using a multiple-infection epidemiological model and quantified as reductions in disability-adjusted life years (DALYs). Labor costs are estimated by modeling the maintenance workload for the ovitrap network as a Vehicle Routing Problem (VRP). Initially applied in Singapore, our analytical

framework is further extended through a simulation study evaluating the prospective cost-effectiveness of implementing a city-wide ovitrap network in Bangkok, Thailand.

2 - Enhancing Mental Health Service Accessibility on University Campuses: the Impact of Tragic Events

Xiaoyan (Amber) Liu, Santa Clara University, Santa Clara, CA, United States, Wilson Lin, Hussein El Haj

Student mental health has emerged as a pressing concern in higher education in recent years. In this paper, we examine how universities can enhance the accessibility of mental health counseling services on campus in light of the potential occurrence of tragic events. Leveraging a comprehensive dataset from the Counseling and Psychological Service (CAPS) center of a US university, we address this issue by first rigorously estimating the impact of a tragic event on the demand for mental health counseling services. We find that demand increases significantly, with individuals being more likely to make an appointment in the immediate days after the tragic event. Second, we propose a multi-period optimization model to allocate mental health clinicians in a rolling-horizon approach given a budget constraint, strategically balancing the trade-off between meeting immediate demand and anticipating future demand. We further show that properly accounting for the demand surge in the optimization framework can yield substantial gains compared to less sophisticated benchmarks.

3 - Hospital-Physician Integration and Quality Outcomes: the Role of Financial Incentives Faced by U.S. Hospitals

Jingyun Su, University of Miami, Coral Gables, FL, United States, Jong Myeong Lim, Masoud Kamalahmadi

The healthcare industry has witnessed a growing trend toward vertical integration between hospitals and physicians. Existing research presents mixed evidence regarding its impact on costs and outcomes. This study fills in the gap of examining relationship between hospital-physician vertical integration and performance under financial incentive programs.

4 - Regression Modeling in Healthcare: Group Fairness and Maximizing the Good

Margret Bjarnadottir, University of Maryland, College Park, MD, United States, Courtney Paulson, Daniel Smolyak

In many healthcare settings, it is both critical to consider fairness when building analytical applications but also uniquely unacceptable to lower model performance for one group to match that of another (e.g. fairness cannot be achieved by lowering the diagnostic ability of a model for one group to match that of another and lose overall diagnostic power). We therefore suggest a new modeling framework, FAIR, to maximize performance across imbalanced groups. This efficient approach overcomes many of the limitations in current approaches and manages to balance learning from other groups with tailoring prediction to the small focal group(s). FAIR has an added advantage in that it still allows for model interpretability in research and clinical settings. We demonstrate its usefulness with numerical and health data experiments.

TA42

Building B Level 3 B302

Organ Transplantation

Invited Session

Health Applications Society

Chair: Paola Martin, Indiana University, Bloomington, IN, United States

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

1 - The Impact of List Diving on Post-Transplant Outcomes

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

Transplant programs frequently turn down deceased-donor kidney offers for the top-ranked potential transplant recipients (PTRs) at their centers and utilize the kidneys for lower-ranked PTRs. This practice is known as list diving. In a retrospective study of 29,932 first-time kidney-only transplants from 18,956 adult donors performed during 2015-2018, we found 6,680 (22.32%) list-diving transplants. PTRs in the US are ranked based on an evidence- and consensus-based policy that reflects the collective judgment of stakeholders in the kidney transplant ecosystem. When transplant programs find it necessary to deviate from the national ranking (either because of PTR unavailability or match-specific reasons), it is expected that their decisions will benefit the recipients they select. However, upon accounting for treatment endogeneity, we find that list-diving recipients' 1-year survival neither worsens nor improves. However, list diving worsens the 1-year survival for transplants from relatively good quality donors (KDPI from 0.2 to 0.35). Conditional on survival, list-diving recipients' are more likely to achieve lower 1-year post-transplant estimated Glomerular Filtration Rate (eGFR-1), which is an indicator of kidney function. The latter is also associated with long-term transplant survival. Additionally, list diving worsens the eGFR-1 for transplants from good quality kidneys.

2 - Value Destruction and Recovery from Supply Chain Disruptions

Yue Tang, Emory University, Atlanta, GA, United States, Nikolay Osadchiy

We study the effect of disruptions in firms' supply chain and internal operations on their market value. We propose two novel empirically grounded metrics to characterize a firm's resiliency to disruptions in its own operations or operations of its supply chain partners: maximum value destruction (MVD) and time to recover market value (TMRV). Applying the event study framework to the disruption data from 2003 to 2019, we show that disrupted firms on average experience 2.1% MVD and need 53.5 days to recover market value. Furthermore, disruptions propagate along supply chains, with significant effects both upstream and downstream and with the downstream supply-disrupted side experiencing higher MVD and longer TTRMV. The effects are mitigated by longer buyer-supplier relationships and customer diversification, but not by multisourcing. A one-year increase in relationship length reduces TTRMV by 8.49% (4.54 days) for disrupted firms, 15.26% (8.13 days) for their customers, and 8.39% (4.01 days) for their suppliers. For suppliers, an additional customer within the same industry as the disrupted firm reduces TTRMV by 10% (6.96 days). In the process, we develop a replicable, shareable, and adjustable fine-tuned large language model-based approach to identifying supply chain disruptions from firm press releases.

3 - Evaluating the Impact of Novel Organ Preservation Technologies on Transplant Access

Ting-Tsen Lin, University of Minnesota, Minneapolis, MN, United States, Saumya Sinha

This project investigates the impact of technological advances on access to organ transplantation. Specifically, we consider novel technologies that preserve organ function for extended periods, enabling organs to be transported over longer distances, improving the quality of transplanted organs, and making certain marginal organs viable for transplantation. We study how such technologies can be efficiently implemented to maximize benefits for transplant recipients and, more broadly, enhance the overall transplant system. The system is modeled as a steady-state framework, formulated as a mixed-integer linear program (MILP). We present both analytical and numerical results.

4 - Targeted Priority Mechanisms in Organ Transplantation

Ruo Chen Wang, Virginia Tech, Blacksburg, VA, United States, Sait Tunc, Burhaneddin Sandikci, Bekir Tanriover, Mark Ellis

The persistent imbalance between organ supply and demand poses a significant challenge for the life-saving treatment of transplantation. This study introduces innovative targeted priority mechanisms inspired by the Eurotransplant Senior Program (ESP), designed to bridge the supply-demand gap while optimizing the matching between organs and recipients without mandating offer acceptance. These mechanisms diverge from traditional prioritization by offering voluntary participation and granting explicit priority to disadvantaged patients for a specified range of organs provided they restrict their offer acceptance to this set. Using a comprehensive queueing model, our study analyzes the strategic decisions of waitlisted candidates under targeted priority mechanisms in equilibrium, establishing the optimal program design. Our findings reveal the capacity of these mechanisms to separate organ allocation across different patient classes, improving overall system efficiency. Our evaluation of these mechanisms' impact on social welfare underscores their potential to improve overall welfare without adversely affecting any patient group. The practical applicability and impact of these mechanisms are demonstrated through a case study within the U.S. kidney allocation system using a clinically detailed simulation model, targeting elderly

patients aged 65 and above. Simulation outcomes reveal a significant participation rate even when the program offers priority over kidneys that are conventionally perceived as marginal. Furthermore, our findings highlight the mechanisms' potential to increase yearly transplants by 1,000 and prevent up to 200 waitlist deaths annually. Improvements in the outcomes of even nonelderly patients highlight the mechanisms' potential to comprehensively address organ underutilization and improve organ-recipient matching in a real-world context.

TA43

Building B Level 3 B303

Healthcare Operations

Invited Session

MSOM: Healthcare

Chair: Huiyin Ouyang, The University of Hong Kong, Hong Kong, Hong Kong

1 - Optimization of Cancer Screening Strategy Using Statistical Control Chart Techniques

Zheng Zhang, Zhejiang University, Hangzhou, China, People's Republic of

This study proposes a two-phase optimization strategy for cancer screening, accounting for binary latent states and random observations. We utilize an Exponentially Weighted Moving Average (EWMA) chart, a tool from statistical quality control, to predict cancer risk and decide whether to proceed with confirmatory testing based on the results of preliminary tests. The problem is formulated as a specialized simulation optimization model, employing decomposition and reformulation techniques to enhance tractability. Our analysis reveals that the model possesses several advantageous properties that significantly improve computational efficiency, facilitating practical implementation in clinical settings. This approach aims to refine decision-making in cancer screening, potentially leading to earlier detection and improved patient outcomes.

2 - The Impact of Physician Familiarity on Emergency Department Productivity

Yiran Zhang, The University of Hong Kong, Hong Kong, Hong Kong, Huiyin Ouyang, Zhankun Sun

Emergency departments (EDs) frequently face challenges with overcrowding, resource limitations, and inefficiencies. Extensive research aims to enhance ED operational performance by understanding physician practices, particularly focusing on physician productivity and patient handoffs. Our working paper examines the impact of physician familiarity on productivity, handoff practices, and patient outcomes, building on existing literature in worker familiarity and productivity. Pairwise physician familiarity is quantified by the number of shared shifts with active interactions. Physician productivity is measured by new patient pickups and instantaneous service rate. Using two years of data from a Canadian teaching hospital, we aim to provide insights for improving ED efficiency.

3 - Feature-Based Optimization of Resource Allocation in Healthcare

Ya-Tang Chuang, National Cheng Kung University, Tainan, Taiwan

Efficient resource allocation is a critical challenge in healthcare systems, particularly under conditions of limited capacity and increasing demand. This study proposes a feature-based optimization framework designed to enhance decision-making by leveraging patient-specific information to improve the allocation of healthcare resources. We model the problem as a Markov decision process (MDP), capturing patient heterogeneity through clinical and non-clinical characteristics. To address the computational challenges associated with high-dimensional state spaces, we propose a priority index (PI) that prioritizes patients based on a cost-benefit analysis between treat-or-wait decisions. A closed-form approximation of the index is derived to enhance computational efficiency, and a Lagrangian relaxation approach is introduced to establish a computable lower bound for performance benchmarking. Numerical experiments demonstrate that the proposed PI achieves near-optimal performance and consistently outperforms benchmark policies, particularly when accounting for patient readmissions. The results underscore the value of efficient resource allocation in enhancing patient flow, advancing equity, and mitigating capacity strain in constrained healthcare environments.

TA44

Building B Level 3 B304

Behavioral and Data-Driven Approaches to Optimizing Healthcare Operations

Invited Session

MSOM: Healthcare

Chair: Galit Yom-Tov, Technion - Israel Institute of Technology, Haifa, Israel

1 - Designing Doctor-Patient Shared Decision-Making Processes Under Bounded Rationality

Feray Tuncalp, Bilkent University, Ankara, Turkey, Rouba Ibrahim, Song-Hee Kim, Jordan Tong

Shared decision-making (SDM) processes, in which doctors and patients work together to choose among treatment options, have gained increasing support from clinicians and governments. Despite the recognized benefits and limitations of these processes, clear guidelines on when and how to implement them are lacking. We develop a stylized analytical model to derive a set of results characterizing when and how to personalize the treatment decision-making process between doctors and patients, taking into account the bounded rationality of doctors and patients. For example, we find that personalizing treatment to account for individual patient preferences is beneficial when treatments have a large expected medical difference across patient types relative to the degree of patient decision error. We also find that the benefits of personalizing treatment to account for individual medical prognoses depends on the degree of patient decision participation. With patient participation, prognosis personalization can even be detrimental when patients strongly prefer one treatment over the other. Challenging common medical belief that advocates for uniform increases in personalization, we show that limitations in medical prognoses accuracy and human cognition imply the existence of trade-offs between personalization and standardization on multiple dimensions within doctor-patient SDM processes. Our results prescribe whether and how to target personalization efforts based on environmental factors, doctor, and patient characteristics.

2 - Justice in Patient Routing

Galit Yom-Tov, Technion - Israel Institute of Technology, Haifa, Israel, Anat Rafaeli

Fairness is critical in the design of queuing and routing systems. We study how patient routing policies affect fairness perceptions and medical personnel behavior. In a field experiment, we examine a new routing policy implemented at a medium-sized hospital, designed to improve fairness in routing patients from the Emergency Department (ED) to three Internal Wards (IWs). This change represents a shift from a pooled queue system to dedicated queues. We assess its impact on fairness, operational performance, and quality-of-care metrics, including patient length of stay (LOS) in both the ED and the wards, as well as the likelihood of readmission.

We find that the new routing policy had positive effects on hospital operations. It reduced overall hospital LOS by approximately one day (14.7%) and decreased ED LOS for patients awaiting IW transfer by 2.2 hours (17%). Importantly, these reductions did not increase the probability of patient readmission, indicating that the policy is a safe and effective way to improve hospital efficiency.

To understand the underlying mechanisms, we conducted online experiments. These reveal that workload imbalance among staff affects fairness perceptions, motivation, and emotional responses such as anger. The dedicated queue policy is perceived as a fairer approach and helps mitigate workload imbalances between wards.

3 - Improving Provider Regionalization: Online Optimization for Patient-Physician Assignment

Angela Lin, Massachusetts Institute of Technology, Cambridge, MA, United States, Dessislava Pachamanova, Georgia Perakis

The practice of localizing physicians to a concentrated geographic region within a large hospital is known as provider regionalization and has been shown to result in a host of benefits.

In this work, we formulate a multi-period stochastic optimization model and design an online algorithm for assigning patients to physicians (after the patient has already been assigned to a unit in the hospital). Our approach accounts for uncertain patient arrivals by simulating future patient arrival sequences using past data distributions, and we propose a novel and realistic objective that is appropriate for the hospital's goals. To make our formulation tractable, we reformulate the problem to eliminate variables, simplify from a nested T-stage stochastic problem to a two-stage problem, relax the integrality of the second-stage variables, and use an enhanced sample average approximation. We also iteratively engage with hospital administrators and physicians to define objectives and incorporate constraints that reflect hospital priorities, and evaluate tradeoffs.

We have an ongoing pilot for our patient-physician assignment framework and have been coordinating with hospital leadership to help change existing practices and workflows. So far, in our pilot, we have seen a 300% decrease in daily physician travel time when using our assignment framework over manual assignment, and additionally, we have saved each assigning physician 3-4 hours of time per week.

Our framework not only saves physician travel time by improving the quality of the assignments, but also saves valuable physician time by eliminating the process of physicians making the assignments themselves.

4 - Advice Provision in the Pandemic: the Impact of Information Granularity on Social Protection

Jingjing Weng, Temple University, Philadelphia, PA, United States, Guangwen Kong, Yang Zhang

This paper investigates the strategies for providing public health advice during a pandemic, such as COVID-19 or a potential future disease, focusing on the granularity of information given by a social planner. We compare two main information provision strategies: aggregated and disaggregated information provision. Under an aggregated information provision strategy, the social planner announces a weighted average of the disease burden across the entire society. Under a disaggregated information provision strategy, the social planner specifies disease burdens for different social groups. We identify conditions where an aggregated information provision is more effective than a disaggregated one. We further extend the analysis of the granularity of information in multi-period scenarios with single- and multi-dimensional information. Our findings show that more granular information does not necessarily lead to better pandemic management and offer insights for designing effective public health information strategies.

5 - Data-Driven Queueing Simulation Using Simlq

Opher Baron, University of Toronto/ SiMLQ, Toronto, ON, Canada, Dmitry Krass, Arik Senderovich

We present a data-driven approach to queueing simulation using SiMLQ. Starting with typical data from queueing networks, we illustrate how such data can be structured as an event log. Leveraging techniques from queue mining and machine learning, we demonstrate how to transform these logs into actionable simulation models. The tale will include a live demonstration of SiMLQ's capabilities across descriptive, predictive, and comparative analytics. We will also discuss how SiMLQ can be applied to prescriptive analytics to support data-informed decision making.

TA46

Building B Level 3 B306

AI and Bayesian Methods for Decision Making

Invited Session

Decision Analysis Society

Chair: Jussi Keppo, National University of Singapore, Singapore, Singapore

Co-Chair: Xinyu Hou, National University of Singapore, Singapore, Singapore

1 - Regularized Aggregation of Point Predictions from Experts with Different Amounts of past Performance Data

Junnan Wang, INSEAD, Singapore, Singapore, Ville Satopaa

Aggregating point forecasts from multiple experts is a critical challenge, especially in the common context where those experts have contributed a varied number of predictions in the past. Consider the scenario where one expert has provided just two forecasts but with perfect accuracy, and another has offered a hundred predictions with lower accuracy. This situation presents a challenge in achieving accuracy while managing the risks associated with less data. To address this, our work introduces a novel aggregation method that balances these considerations. Specifically, our aggregator factors in their unique prediction histories, and appropriately shrinks each expert's weight toward equal weight, depending on the expert's skill and how reliably that skill can be estimated. In our empirical study, we apply our aggregator to the data from the European Central Bank's Survey of Professional Forecasters and observe that it not only outperforms traditional aggregation methods, such as the simple average methods, but also improves the accuracy of performance-weighted methods and performance-based crowd-selection methods.

2 - Diversified Learning: Bayesian Control with Multiple Biased Information Sources

Xinyuan Zhang, Sauder School of Business, University of British Columbia, Vancouver, BC, Canada, Jussi Keppo, Michael Jong Kim

We consider a decision-maker (DM) who can acquire signals from multiple biased information sources to learn about a hidden state before making an earning decision. Unbiased signals are also available but come at a higher acquisition cost. The DM jointly optimizes both learning (information acquisition) and earning decisions to minimize expected loss. This problem is motivated by applications such as medical diagnostics, revenue management and financial investments, where decisions often rely on multiple biased information sources. In contrast to existing literature that primarily focuses on stopping problems with unbiased information, we develop a Bayesian decision framework that accommodates general earning decisions, and models multi-source learning using a hierarchical Bayesian network, explicitly capturing intrinsic biases. We fully solve the model and explicitly characterize the optimal acquisition policy: For small budgets, cost-effectiveness dominates, prioritizing cheap biased sources. As budgets grow, diversification across all biased sources becomes optimal as it mitigates risks from biases. With large budgets, acquisition shifts to costly unbiased sources as biased signals offer diminishing value due to their limited accuracy. We also illustrate the model through two use cases: a disease nowcasting example using real data that shows how budget allocation shifts with bias, cost, and precision, and a demand forecasting example that demonstrates the performance advantage of the proposed policy (learning algorithm) over benchmarks in simulation

3 - Risk-Averse Decision-Making with Endogenous Regressors

Xinyu Hou, National University of Singapore, Singapore, Singapore, Jussi Keppo, Ian Zhu

Data-driven optimization often follows an estimate-then-optimize paradigm, but endogeneity introduces bias that can degrade decision quality. We propose a Bayesian framework that compares Bayesian linear regression (which is biased under endogeneity) with the corresponding Bayesian two-stage least squares (2SLS), which leverages instruments to correct the bias. While 2SLS addresses endogeneity, it also increases the posterior variance. This raises a critical question: Should risk-averse decision-makers always prefer unbiased estimators?

We examine this bias-variance tradeoff and show that unbiased estimators like 2SLS do not necessarily lead to better outcomes when both estimation uncertainty and downstream decision performance are considered. Our results guide estimation method selection by linking endogeneity correction not only to statistical consistency but also to prescriptive performance under risk, offering practical insights for robust decision-making under imperfect identification.

4 - Pasta: a Unified Framework for Offline Assortment Optimization

Juncheng Dong, Duke University, Chapel Hill, NC, United States, Weibin Mo, Zhengling Qi, Cong Shi, Ethan Fang, Vahid Tarokh

We study a broad class of assortment optimization problems in an offline and data-driven setting. In such problems, a firm lacks prior knowledge of the underlying choice model, and aims to determine an optimal assortment based on historical customer choice data. The combinatorial nature of assortment optimization often results in

insufficient data coverage, posing a significant challenge in designing provably effective solutions. To address this, we introduce a novel Pessimistic Assortment Optimization (PASTA) framework that leverages the principle of pessimism to achieve optimal expected revenue under general choice models. Notably, PASTA requires only that the offline data distribution contains an optimal assortment, rather than providing the full coverage of all feasible assortments. Theoretically, we establish the first finite-sample regret bounds for offline assortment optimization across several widely used choice models, including the multinomial logit and nested logit models. Additionally, we derive a minimax regret lower bound, proving that PASTA is minimax optimal in terms of sample and model complexity. Numerical experiments further demonstrate that our method outperforms existing baseline approaches.

TA47

Building B Level 3 B307

Learning Enhanced Optimization

Contributed Session

Contributed

Chair: Jack Zhang, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Integrated Offline and Online Learning to Solve a Large Class of Scheduling Problems

Anbang Liu, Tsinghua University, Beijing, China, People's Republic of, Zhi-Long Chen, Jinyang Jiang, Xi Chen

We propose a machine learning approach to efficiently generate high-quality solutions for single-machine scheduling problems with a non-decreasing min-sum objective function with or without release times. Our approach is novel in three major aspects. First, through exploiting the time-index formulation and Deep Neural Network (DNN) model, our approach provides a unified solution method for the entire class of the aforementioned problems. The second novel aspect lies in the training of our DNN model. In view of the NP-hard nature of the problems, labels (i.e., optimal solutions) are hard to generate for supervised training. To overcome this difficulty, we generate a set of special instances, for which optimal solutions can be found with little computational effort, to train the DNN model in a supervised manner offline. The third novel idea is that we develop an online single-instance learning approach to fine tune the offline trained parameters in the DNN for a given online instance, with the goal of generating an improved solution for the given instance. Computational results show that our approach can quickly generate near-optimal solutions for a variety of single-machine scheduling min-sum problems with up to 1000 jobs.

2 - Automatic Change-Point Detection Using Piecewise Support Vector Quantile Regression

Sheng-Yu Tsai, Department of Information Management, National Chi Nan University, Nantou, Taiwan, Jing-Rung Yu, Chun-Yen Tsai, Donald Lien

Pattern change detection (PCD) in time series data is essential for identifying structural shifts, enabling accurate forecasting and informed decision-making across domains such as finance, healthcare, IoT, and manufacturing. This study proposes a piecewise support vector quantile regression (PSVQR) model to achieve automatic change-point detection and accurate interval prediction. The PSVQR model combines four key components: it employs the quantile loss function to estimate conditional quantiles, generating prediction intervals that account for non-constant variance and data heterogeneity; it incorporates a piecewise structure formulated through multi-objective programming to capture structural changes by segmenting data into distinct regions; it enables automatic detection of multiple change-points by treating their number as an optimization objective; and it utilizes feature selection to refine input variable selection and improve model parsimony. Overall, the PSVQR framework offers a flexible, adaptive, and computationally efficient solution that outperforms conventional methods in handling complex datasets with structural shifts.

3 - Formulation and Regret Analysis of a Delivery Reward Maximization Problem Using Contextual Multi-Armed Bandits with Stochastic Costs

Shinnosuke Uba, The University of Osaka, Suita, Japan, Yutaro Yamaguchi

In food delivery, a courier moves across multiple locations, deciding whether to accept or reject delivery requests to maximize the cumulative reward within a limited time. Since each location offers different reward efficiencies and waiting times, efficient strategies must balance exploration and exploitation under uncertainty. We model this problem as a contextual multi-armed bandit problem with stochastic costs. Each location is associated with a set of actions: either accepting a request at the current location, or taking an action that moves to a (possibly the same) location to wait for the next request. In this model, the current location corresponds to the context. To solve this problem, we propose an UCB-based algorithm that leverages the structure of the problem. Under several technical assumptions, such as the destination of each request being drawn uniformly at random from all cities, we prove that the expected regret of this algorithm, compared to an optimal algorithm with full knowledge of the environment, increases at the best possible logarithmic rate with respect to the number of decisions. Due to the difference in the number of independent targets for exploration, our method achieves a better regret bound in terms of the constant factor compared to directly applying the previously proposed method for the contextual multi-armed bandit problem with stochastic costs, where both the context and the cost distribution are of a single type.

4 - Neural Approximate Optimization for Real-TIME Dispatching of Renewable Energy Sources with Battery Energy Storage

Minseo Kim, Yonsei University, Seoul, Korea, Republic of, Soongeol Kwon

We address the problem of real-time dispatching for renewable energy sources integrated with battery energy storage. Given day-ahead commitments, i.e., the amount of electricity committed for sale ahead of operations, real-time dispatching is designed to maximize profit by determining the amount of electricity to sell in the real-time market while using battery storage to compensate for generation mismatches due to uncertainty. This problem is formulated as a multi-stage stochastic program (MSSP), where non-anticipativity under evolving uncertainty in renewable energy generation and electricity prices makes it challenging to ensure feasibility with time-dependent constraints and achieve profit-maximizing optimality. To overcome these challenges, we propose a neural approximate optimization framework that approximates the decision-making process of the MSSP in a computationally efficient and scalable manner. The framework reduces the dimensionality of the decision space through variable reduction, resulting in a single net dispatch adjustment variable. Input sequences are constructed using only realized and available information up to the current period. A Transformer encoder is employed to capture temporal dependencies, with a feasibility-aware module embedded to enforce both equality and inequality constraints through variable reconstruction and bound-based projection. The model is trained using multi-step-ahead targets derived from optimal solutions, along with a profit-oriented loss function that directly aligns learning objectives with profit maximization. As a result, the proposed framework yields feasible and profitable dispatching decisions without requiring explicit optimization during real-time operation.

5 - Heatflow: a Non-Autoregressive Solution to the Workflow Scheduling Problem

Vivek Anand Rajopanth, University of Illinois Urbana-Champaign, Urbana, IL, United States, Rakesh Nagi, Asser Tantawi

Efficiently orchestrating large batches of data and compute-intensive workflows is a central challenge for modern cloud platforms. Existing production schedulers rely on hand crafted heuristics that trade solution quality for speed, whereas exact approaches such as Mixed-Integer Linear Programming (MILP) can deliver optimal schedules that are prohibitively slow for use in practice. We introduce HeatFlow, the first non-autoregressive Graph Neural Network (GNN) framework that is MILP-aided yet inference-time efficient for the workflow scheduling problem. HeatFlow consists of GNNs that predict an ordering of operations along with an efficient allocation of operations to resources. The resulting “heat maps” are decoded using a lightweight heuristic that guarantees feasibility. Training labels are harvested offline from a MILP that accounts for all relevant constraints in the cloud environment. Experiments on three diverse, real-world and simulated workload suites running on a simulated KWOK cluster controlled using Kubernetes demonstrate that HeatFlow produces solutions that outperform traditional baselines in terms of optimality gaps with minimal computational time. These results highlight the promise of learning-augmented optimization for production cloud schedulers and for other similar combinatorial optimization problems.

6 - Discovering Closed-Form Solutions to Quadratic Programs with Linear Constraints with Neural Networks

Fuat Can Beylunioglu, University of Waterloo, Waterloo, ON, Canada, P. Robert Duimering, Mehrdad Pirnia

Mathematical programming has primarily been a study of numerical optimization where the solution is obtained following a procedure recursively until convergence. Recent applications of Neural Networks (NN) as surrogate models have challenged this view by treating optimization problems as approximating unobserved functions that map input parameters to optimal solutions, similar to closed-form solutions. Drawing on insights from multiparametric programming (MP), which explores piecewise-linear (PWL) properties of quadratic programs with linear constraints (QP), we propose a closed-form NN (CF-NN) model and a learning via discovery algorithm that

learns an exact representation of the solution function without training, by applying algebraic operations to derive all model weights from problem coefficients without the need for training data. Our proposed learning algorithm begins with an optimal solution to initiate the NN model, then discovers each linear segment of the PWL solution function and expands the model parameters iteratively without using solvers. The proposed procedure yields a NN model that acts as the closed-form solution to QP with linear constraints, ensuring that optimal solutions for the discovered regions are both feasible and optimal. It is shown that the proposed CF-NN model offer significant advantages. The model takes seconds to learn the solution function and can seamlessly generalize outside the training distribution. Results on IEEE DC-OPF test cases significantly outperforms traditional deep NN training methods and a commercial solver, achieving optimal solutions with around $1E-16$ KKT optimality violations in all discovered critical regions.

7 - Using Deep Learning to Approximate Feasible Solutions for General Scheduling Assignment Problems

Constanza Lorca, University of Virginia, Charlottesville, VA, United States, Robert Riggs

In some real-world scheduling environments, the availability of resources can vary unpredictably, requiring schedulers to make quick decisions that are not always optimal. Solving these problems dynamically frequently limits the ability to guarantee optimality, showing the need for fast approximation methods.

To tackle the challenge of creating feasible solutions in a timely manner, we propose a neural network framework that learns to approximate feasible near optimal solutions for an assignment decision problem. This approach implicitly incorporates constraints during the learning process, to enable fast and valid assignment predictions, based on optimal solutions generated by a mathematical programming solver.

Our model learns to predict assignment probabilities, by incorporating a masked SoftMax layer into the neural network model, that enforces constraints and only allows for valid assignments. It is trained on the solutions generated by a MILP solver across diverse generated instances, with varying problem sizes and constraints, which ensures generalization and scalability.

This work aims to integrate operations research and deep learning, by demonstrating how neural network models can be used to approximate combinatorial problem solutions while ensuring feasibility and reducing computation time. This framework provides an efficient decision-support tool to generate solutions to reallocate tasks in dynamic scheduling environments.

8 - Approximate Solution of Infinite-Horizon Risk-Sensitive Markov Decision Processes

Jack Zhang, Massachusetts Institute of Technology, Cambridge, MA, United States, M. Soheil Hemmati, Saumya Sinha

Infinite horizon risk-sensitive Markov decision processes (MDPs) with the discounted cost criterion are challenging to solve because the optimal policy may be non-stationary. Existing solution methods reformulate the problem as a continuous-state (risk-neutral) MDP and solve it using state-discretization or value function approximation. In this paper, we present approximate variants of value iteration, policy iteration, and linear programming approaches that can directly solve the risk-sensitive MDP without the need for reformulation. We show that the approximate value functions obtained by value iteration and linear programming converge to the optimal value function, and the policies generated by the policy iteration algorithm converge in value to the optimal policy. Importantly, in addition to proving asymptotic convergence, we provide explicit stopping conditions to achieve arbitrarily close approximations to the optimal value function.

TA48

Building B Level 3 B308

Advances in Decision Making: Theory and Applications

Invited Session

Decision Analysis Society

Chair: Hao Wang, University of Science and Technology of China, Hefei, N/A

Co-Chair: Guangyan Gan, NTU, Singapore, Singapore

1 - Online Emergency Resource Allocation in Humanitarian and Health Logistics

Hao Wang, University of Science and Technology of China, Hefei, China, People's Republic of

Resource allocation is a critical challenge in humanitarian and health logistics (HHL) such as allocating scarce resources to patients or disaster-affected populations. Given the dynamic and uncertain nature of such systems, where decisions must be made with limited information about future supply or demand and without the possibility of recourse, these problems in HHL are referred to as online emergency resource allocation. To address this challenge, we propose an online resource allocation model for emergency management and an adaptive algorithm designed to minimize the performance gap relative to static offline benchmarks. We prove that our algorithm achieves a competitive ratio of $1/8$ and establish a hardness result demonstrating that no algorithm can attain a ratio better than $1/2$. To validate our approach, we conduct a case study using data of in-kind donations following the 2023 Jishishan earthquake. The results confirm the feasibility of our model, demonstrate the effectiveness of our solution, and illustrate how our algorithm can be applied in practice.

2 - Fostering OR Hindering Sales? Overlapping Influencer Followers in Live Streaming E-Commerce

Jingyun Hu, Clemson University, Pendleton, SC, United States

In social media, consumers often follow multiple influencers and may encounter advertising messages for the same product from different influencers. When firms switch influencers to promote their products, an overlap in followers between the previous and current influencers may diminish the returns from the promotion but can also provide additional opportunities to sell to followers of the first influencer who have not yet purchased. This study develops a formal model to examine the impact of overlapping followers on a firm's incremental sales when engaging a new influencer. We generate and empirically test predictions using live streaming e-commerce data from Douyin (the Chinese sister app of TikTok). By utilizing Natural Language Processing (NLP) to estimate the follower overlap between influencers and their respective product demands, we find that more follower overlap often reduces profits but can increase profits if these overlapping followers have a disproportionately high demand for the firm's product. For example, overlap can be beneficial in cases where the firm did not advertise too heavily through the first influencer and can still generate significant sales by re-engaging these consumers through the second influencer. Heterogeneity analysis suggests that the effects of follower overlap vary substantially across different product categories.

3 - Dynamic Pricing and Resource Relocation in a Large Closed Network

Shubo Sun, Miami Herbert Business School, Miami, FL, United States, Ganggang Xu, Xu Sun

We study dynamic pricing and resource relocation in a closed service network with stochastic customer arrivals and travel times. Motivated by shared mobility and micromobility platforms, we consider a system with $n + n_0$ supply units distributed across multiple locations, where n represents the ideal server size under instantaneous relocation and n_0 provides safety staffing to buffer against supply unavailability caused by relocation delays. We propose a state-dependent policy that integrates real-time pricing with adaptive relocation and prove that it achieves a per-edge revenue loss of $O(n^{-(2/3)} \log(n))$ relative to a fluid model benchmark for a given network

structure. Our analysis reveals that this optimality gap is attainable with the additional safety staffing n_0 , whose magnitude depends on the network topology and relocation speeds. This quantifies the operational cost of travel-time stochasticity and demonstrates that pricing alone cannot ensure efficiency in large-scale stochastic networks. Our framework offers a tractable and provably effective approach for demand-supply coordination in modern service networks.

4 - Modeling the Epidemiology Dynamics of Respiratory Syncytial Virus Infections in the United States Under Novel Pharmaceutical-Based Interventions

Xinmeng Zhao, Centers for Disease Control and Prevention, Atlanta, GA, United States, Brian Gurbaxani, Michael Melgar, Ismael Ortega-Sanchez, Ian Plumb, Ryan Wiegand, Phillip Salvatore

Respiratory syncytial virus (RSV) is the leading cause of hospitalization among infants and is associated with 100,000–150,000 admissions annually in U.S. adults aged ≥ 60 years. The recent implementation of novel immunizations—including maternal vaccines, long-acting monoclonal antibodies for infants, and vaccines for pregnant women and other adults—demands a framework capable of quantifying their population-level impact, including both direct protection and indirect effects. While previous mathematical models have described RSV transmission, few have evaluated the impacts of interventions using economic and morbidity metrics, and none have evaluated multiple simultaneous interventions. For the first time, we report the development process of a novel, multi-intervention model for rapid estimation of disease burden to evaluate intervention guidelines. We developed a deterministic susceptible-exposed-infected-recovered (SEIR) model stratified by four RSV infection histories (maternal, primary, secondary, repeat), 32 age classes (monthly intervals through 24 months; broader bands thereafter), and two clinical-risk groups. Each history tier comprises asymptomatic, symptomatic, hospitalized, and recovered compartments, with births, ageing, natural mortality, and waning immunity explicitly represented. Implementation of this model is customizable to end-user localities, burden, and seasonality through adjustment of calibration targets to population demographics. This age- and risk-resolved platform provides a tool for assessing current and future RSV intervention guidelines across diverse subpopulations.

TA49

Building B Level 3 B309

JM Candidates - Technology, Society and Policy

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Alp Sungu, Wharton, Philadelphia, PA, United States

1 - Performance Gains, Motivational Pains: Evidence from AI Adoption in Turkish Schools

Benjamin Lira Luttgies, University of Pennsylvania, Philadelphia, PA, United States, Alp Sungu, Angela Duckworth

While educators express growing concern over generative AI's role in teaching, empirical evidence of its classroom impact remains scarce. We conducted a large-scale randomized field experiment in Turkish private schools to evaluate whether access to an AI-based teaching assistant—and behavioral nudges to encourage its use—could enhance educational outcomes. Teachers were randomly assigned at the school-department level to three conditions: (1) control group receiving standard professional development, (2) AI-access group trained on a custom teaching tool, and (3) AI-exposure group receiving the same training plus weekly engagement reminders.

Both AI conditions significantly improved student performance, with the AI-access group showing the largest gains. However, this academic success came at a cost: the AI-access group experienced marginal declines in both teacher and student motivation, alongside more negative attitudes toward AI. In contrast, the AI-exposure group—encouraged to engage more consistently with the tool—achieved smaller student gains but no motivational penalties.

We explore treatment heterogeneity and conduct exploratory language analysis of chatbot logs to better understand usage patterns.

Our findings highlight a critical tension in educational AI adoption: the conditions producing the strongest immediate academic improvements may simultaneously erode the motivational foundations necessary for sustained success. These results suggest that successful AI integration requires not merely access but carefully structured implementation that balances performance gains with the preservation of teacher autonomy and student engagement.

2 - Spatial Matching Under Multihoming

Alireza Amaniamedani, London Business School, London, United Kingdom, Ali Aouad, Daniel Freund

Motivated by spatial matching platforms such as ride-hailing, we present a model of two-sided matching platforms in a duopoly market with spatial frictions. The two platforms face disjoint streams of rider requests, but they share the same suppliers, i.e., the suppliers multi-home. Spatial frictions create a need for admission control policies that maintain a buffer of available suppliers to balance matching distances against other costs. We find that a duopoly equilibrium entails at least one platform undercutting, by accepting all requests, to gain market share. When both platforms undercut, the supply buffer is not maintained and inefficiencies ensue. We provide a classifier of market conditions that delineates the two regimes and allows us to characterize the prices of anarchy and stability in these settings.

3 - Large-Scale Optimal Contextual Pricing for Demand Response

Mohammad Mehrabi, Stanford University, Stanford, CA, United States, Omer Karaduman, Stefan Wager

The time at which renewable (e.g., solar or wind) energy resources produce electricity cannot generally be controlled. In many settings, consumers have some flexibility in their energy consumption needs, and there is growing interest in demand-response programs that leverage this flexibility to shift energy consumption to better match renewable production—thus enabling more efficient utilization of these resources.

We study optimal demand response in a model where consumers operate home energy management systems (HEMS) that can compute the “indifference set” of energy-consumption profiles meeting pre-specified consumer objectives, receive demand-response signals from the grid, and control consumer devices within the indifference set. For example, if a consumer asks for the indoor temperature to remain between certain upper and lower bounds, a HEMS could time the use of air conditioning or heating to align with high renewable production when possible.

Here, we show that while price-based mechanisms do not generally achieve optimal demand response—i.e., dynamic pricing cannot induce HEMS to choose optimal demand-consumption profiles within the available indifference sets—pricing is asymptotically optimal in a mean-field limit with a growing number of consumers. Furthermore, we show that large-sample optimal prices can be efficiently derived via an algorithm that only requires querying HEMS about their planned consumption schedules under different prices. Finally, we study learning the optimal price passively (when frequent queries to HEMS are not possible) as an online learning problem over a series of days, using only the day's forecasts—i.e., weather, temperature, and weekday conditions—as context, and by observing only consumer consumption profiles.

TA50

Building B Level 3 B310

Empirical Research on Responsible Operations

Invited Session

MSOM: Service Operations

Chair: Amrita Kundu, McDonough School of Business, Georgetown University, Washington, DC, United States

Co-Chair: Kamalini Ramdas, London Business School, London, United Kingdom

1 - Working with AI vs Humans: The Impact on Wellbeing

Shawn Mankad, North Carolina State University, Raleigh, NC, United States, Tim Kraft, Eirini Spiliotopoulou, Yanchong Zheng

The use of AI decision-support tools can benefit operational efficiency. However, it can also impact employee wellbeing. In this paper, we utilize human-subject lab experiments to examine how collaborating with an AI tool (versus a human co-worker) impacts both worker performance and wellbeing.

2 - Sequential Selection of Candidates: An Experimental Investigation

Morvarid Rahmani, Georgia Institute of Technology, Atlanta, GA, United States, Karthik Ramachandran, Chris Green

In many settings, managers face sequential selection decisions, constrained by both the number of selections they can make and the uncertainty about forthcoming candidates. This "sequential selection" problem is exemplified in the United States Army's officer evaluation process, where supervising officers assess subordinates' performance and decide whether to award a "high" rating. Critically, these decisions are further restricted by strict limits on the proportion of evaluated officers who can receive high ratings. This paper employs a theoretical model and develops an experimental framework to study how effectively human subjects select candidates and to explore the factors influencing their decisions.

3 - Measuring the Impact of Solar Lighting on the Lives of the Poor: Implications for Donor Organizations and off-Grid Solar Firms

Bhavani Shanker Uppari, Singapore Management University, Singapore, Singapore, Serguei Netessine, Michael Sakellakis

We randomly assigned solar-lighting kits with varying configurations to a large number of households in Ghanaian villages. We document the impact of these kits on the treatment households, in terms of economic improvement as well as improvements in the education of children and the overall well-being of the household members, vis-à-vis the control households that were not given solar kits. We find that the treatment households decreased the purchase and usage of inefficient and unsustainable energy solutions, they recharged mobile phones at home, and the overall expenditure on medicines decreased. These contribute to increased savings: a treatment household (on avg.) saved 83 GHS per month more than a control household, which amounts to 20% of their earnings. The treated children spent (on avg.) 20 minutes longer studying and scored 13 percentage points higher than the children in the control group. Interestingly, the impact magnitudes were similar across different solar-kit configurations in our experiment. This is because the households that received less and/or dimmer bulbs compensated for their lack of sufficient (brighter) light by purchasing additional USB-compatible light sources that could be recharged at home using their short-term savings. We discuss the implications of this reactionary behavior of consumers for the donor organizations (in terms of how they can optimize their donation resources) and the off-grid solar firms (in terms of designing better payment models).

4 - Retailer's Informational Role in the Durability of Goods in An Informal Circular Economy

Amrita Kundu, McDonough School of Business, Georgetown University, Washington, DC, United States, Erica Plambeck, Qiong Wang

This paper reveals the strategic behaviour of small retailers in providing (mis)information to heterogeneous customers about the durability of goods from competing brands in an informal circular economy. These retailers sell lead-acid batteries for electric three-wheeled vehicles, buy-back batteries at end of life, and thus directly observe the life of batteries of various brands. Most batteries have inefficiently short life, whereas customers want long life. A retailer could guide customers to the longest-lasting brand for which the marginal cost is well justified, to build reputation and attract a stream of repeat customers. To observe how retailers actually behave, we covertly audit 261 retail shop owners. Posing as five different types of customer across 1,122 visits, we learn that retailers provide inconsistent information about battery life and which brand is longest-lasting. For poor customers, a retailer exaggerates the life and sets a higher price (net of used battery trade-in value) for a given brand of battery. More frequently with poor customers, a retailer promotes as longest-lasting the brand that offers him the best incentive. This puts poor customers (the predominant types of customer in the market) at an informational and economic disadvantage. For wealthy customers with multiple vehicles, a retailer sets the lowest prices for a given brand of battery. Manufacturers of long-lasting batteries or policy-makers should establish systems to provide reliable information to customers about battery life. Adoption of long-lasting batteries would reduce the rate of lead emissions from battery recycling and manufacturing, which is crucial for public health.

5 - The Impact of Racial Matching on Foster Care Outcomes

Wani Zhang, London Business School, London, United Kingdom

We investigate the impact of racial matching of foster children and caregivers on placement stability and permanency outcomes using data from the Adoption and Foster Care Analysis and Reporting System (AFCARS) from 2008 to 2021. Utilizing hazard models, we find strong evidence that racial matching significantly reduces the risk of placement transfers, indicating enhanced placement stability. Additionally, racial matching affects permanency outcomes, increasing the hazards of adoption and guardianship while decreasing the hazard of reunification with biological families. Our results are robust to several alternative specifications that account for sources of endogeneity. Our findings highlight the complex and crucial role racial matching plays within the foster care system, suggesting that placement practices promoting racial similarity could substantially improve children's long-term stability and well-being.

TA51

Building B Level 3 B311

Emerging Topics in Analytics Research for Social Good

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: Somya Singhvi, USC Marshall School of Business, Los Angeles, CA, United States

Co-Chair: Divya Singhvi, New York University, New York, NY, United States

1 - Recommendations in High-Stakes Settings

Nikhil Garg, Cornell Tech, New York, NY, United States

Recommendation systems are now used in high-stakes settings, including to help find jobs, schools, and partners. Building public interest recommender systems in such settings bring both individual-level (enabling exploration, diversity, data quality) and societal (fairness, capacity constraints, algorithmic monoculture) challenges. In this talk, I'll discuss our theoretical, empirical, and deployment work in tackling these challenges, including ongoing work on (a) applicant behavior and recommendations for the NYC HS match, (b) a platform to help discharge patients to long-term care facilities, (c) feed ranking algorithms on Bluesky for research paper recommendations.

2 - The Disparate Effect of Disasters on American Supply Chains

Mark Brennan, Rutgers University, Camden, NJ, United States, Jonas Jonasson

Recent disasters across the United States highlight the devastating effect of climate-change. How these disasters impact business continuity is a pressing issue of social equity and public policy. In this work, we investigate the impact of disasters on supply chains. We first estimate the impact—as well as the subsequent rate-of-recovery—of severe floods on different tiers of the supply chain. We then investigate the level of heterogeneity of those effects by sector and along socioeconomic lines. Finally, we explore whether recovery at the retail tier of the supply chain is driven by local recovery at the wholesale tier. This has implications for disaster relief policies and resource allocation.

3 - Exploring Classical Heuristics and Biases in LLM-Assisted Decision Making

Jessica Shi, New York University, New York, NY, United States

As large language models (LLMs) become increasingly embedded in data-centric workflows, their impact on human judgment and decision-making warrants careful scrutiny. This study examines how cognitive biases manifest when individuals use LLM-generated summaries to make decisions in data science contexts. Through a series of controlled experiments grounded in realistic data evaluation tasks, we explore how variations in the presentation of information from LLM-generated summaries influence user decisions. Our results reveal systematic patterns of bias, including framing effects, negativity bias, anchoring, and narrative bias, all of which collectively diminish critical engagement with underlying statistical evidence. Moreover, we find that users' follow-up queries to the LLM often reinforce their initial biases, creating a

feedback loop that exacerbates miscalibrated decision-making. To address this, we design interventions that nudge users toward more reflective questioning strategies, improving deliberative reasoning. These findings underscore the importance of bias-aware LLM interfaces and explanation strategies to foster more calibrated and responsible human-AI collaboration.

4 - A Data-Driven Approach to Improve Artisans Productivity in Distributed Supply Chains: Evidence from Field Implementation

Xinyu Zhang, New York University Stern School of Business, New York, NY, United States, Divya Singhvi, Somya Singhvi

Despite their vital role in rural economy, artisanal supply chains continue to be plagued by low productivity and the highly fragmented nature of their upstream. This study presents research conducted in close collaboration with a leading exporter of handmade rugs in India, aimed at improving artisans' productivity in distributed supply chains. Inspired by the robust empirical evidence that frequent supervisor visits can play a crucial role in improving artisans' productivity, we implement a novel predict-then-optimize framework for optimizing supervisor visits in the supply chain for a set of treatment looms for 25 consecutive weeks. A difference-in-differences (DID) analysis using data from more than 5,700 supervisor visits across 200 looms during the implementation suggests that the implementation is associated with a statistically significant 16.5% increase in weaving rate for treatment looms, highlighting the practical relevance of the work.

5 - To Pivot or Persevere? An Examination of Venture Founder Solution Choice Strategies

Stylianos Kavadias, University of Cambridge, Cambridge, United Kingdom

A recent surge in the strategy and entrepreneurship literature has reopened search and dynamic choice models as a way to capture pivots in entrepreneurial ventures. Motivated by several examples of pivots we have studied in a European business school accelerator, in this study we are introducing a model of venture founder decision making that reflects both the bounded rationality and the imperfect information they are facing when making choices about pivoting or persevering. Within that context, we delineate some of the key parameters that describe the solution landscape and we develop a framework that showcases that in certain circumstances persevering (also viewed as a less flexible choice strategy) ends up being the best choice strategy for the founder.

TA52

Building B Level 3 B312

Equity and Sustainability in Supply Chains

Invited Session

MSOM: Sustainable Operations

Chair: Karthik Murali, Oregon State University, Corvallis, OR, United States

Co-Chair: Aditya Vedantam, University at Buffalo, Buffalo, United States

1 - Enhancing Drug Development Success and Safety Through Attention to Physiological Differences

Nagarajan Sethuraman, University of Kansas, Lawrence, KS, United States, Deepak Jena, Rachna Shah

We examine whether gender diversity in pharmaceutical invention teams affects drug development success by analyzing 422,822 patents. We show that gender-diverse teams create patents more likely to contribute toward approved drugs, with improved safety outcomes.

2 - Recycle and Resale: Can Fashion Brands Have It All?

FAN SI, National University of Singapore, Singapore, Singapore, Gongao (Lucy) Chen

In recent years, many fashion brands have launched trade-in programs and begun selling second-hand products through their official websites. In this paper, we examine the impact of such programs on firms' profitability and their environmental implications. We find that such programs can sometimes reduce a firm's profit. Moreover, different payment schemes provided to second-hand product owners can have different impacts on both the firm's performance and environmental outcomes.

TA53

Building B Level 3 B313

On Search and Choice

Invited Session

Revenue Management

Chair: Aydin Alptekinoglu, Penn State, University Park, PA, United States

1 - Designing Size Inclusive Fashion Assortments

Alireza Alavi, McGill University, Montreal, QC, Canada, Mehmet Gumus, Javad Nasiry

Size unavailability—the absence of apparel in certain sizes—damages both profit and brand image. Plus-size customers face this unavailability far more frequently than others, creating a disparity that has been called "a subtle form of social discrimination." In this research, we address this problem in a static setting. We incorporate customers' reactions to size unavailability into our choice model and develop algorithms to create fashion assortments that are both profitable and size-inclusive.

2 - Dynamic Assortment in Service Systems

Chen-An Lin, Purdue University, West Lafayette, IN, United States, Yuan Guo

This paper examines dynamic assortment optimization in service systems where customer purchasing decisions are influenced by real-time congestion, as seen in food delivery platforms. We address two key questions: (1) How should firms adjust product or service variety when customer choices depend on evolving congestion levels? and (2) How do product features, such as profit margins and service times, shape optimal assortment strategies? We model the system as an M/G/1 queue and develop a dynamic control framework under a general discrete choice model with impatient customers, where the system state reflects congestion and the decision variable is the assortment offered. Each product is characterized by its selling price, production cost, and ex ante service time distribution. We first establish the existence of a concave solution function under a general choice model. For the MNL choice model, we show that the optimal policy follows an adjusted-margin-ordered structure, where product adjusted margins depend on product features and congestion. We find that when product margins increase linearly with service time, the optimal assortment expands in size as congestion rises. For more general relationships between product margins and service times, we find that optimal assortments in service systems can be non-monotonic, reflecting richer trade-offs than those captured by traditional revenue-ordered policies. In a special two-product setting, we provide sufficient conditions for this non-monotonicity. Our results reveal a novel tradeoff consisting of two counteracting forces: while higher congestion may lead to smaller assortments to improve service efficiency, firms may also expand offerings to sustain demand.

3 - Optimal Search for Tiered Product Information

Zhen Liu, Benedictine University Goodwin College of Business, Lisle, IL, United States

In this study, we assume the various ways a consumer can gather information about a product, including paying search costs before purchase, obtaining information through ownership, and participating in trials (free or paid) before making a purchase. Our model incorporates all three methods of information gathering and analyzes the consumer's decision-making process, including the decision to initiate trials without incurring further search costs, the duration of the trial period, and the ultimate decision to purchase the product.

In this study, we model the customer's expected valuation as a continuous-time Brownian motion and formulate the customer's decision-making process as a multiple optimal stopping problem, following the approach of Branco et al. (2012). Our research characterizes the optimal stopping rules for either starting a trial or not, as well as the optimal stopping rules for either purchasing the product or not, taking into account factors such as search costs, attribute informativeness, discount factor, and trial and sales prices. Our findings indicate that the optimal trial period depends on discount factor, trial and sales prices, but is independent of search costs and informativeness. We

also examine the seller's optimal pricing strategies, considering factors such as discount factors, switching costs, and freemium pricing strategies (limited access to attributes with longer trial periods). Our results support the widespread use of fixed-time trial periods in the SaaS industry and provide a basis for marketing managers to make informed pricing decisions.

4 - Shaping Choices in Sequential Search

Aydin Alptekinoglu, Penn State, University Park, PA, United States, Victor Martinez-de-Albeniz

We study how a retailer can shape the consumers' sequential search outcomes by setting product-specific search costs and choosing an assortment. Building on Weitzman (1979) we first characterize when it is optimal for the retailer to induce a search sequence that is "revenue-ordered" (in decreasing order of prices or, equivalently, margins). Then, we explore how much additional expected revenue (or, equivalently, profit) the optimal sequential-search solution brings over the traditional assortment planning solution. In essence, the goal of this analysis is to compare two fundamental ways of framing assortment optimization: One assuming that consumers search and consider products one at a time (sequentially, à la Weitzman, 1979); and the other assuming that consumers consider all available products at once (in parallel), which is more traditional and in line with most models of assortment optimization in the extant literature.

TA54

Building B Level 2 B216

Frontiers of Revenue Management: Theory and Practice

Invited Session

Revenue Management

Chair: Ming Hu, Rotman School of Management, University of Toronto, Toronto, ON, Canada

Co-Chair: Ali Daei Naby, Rotman School of Management, University of Toronto, 468 Roxton Rd., Toronto, ON, M6G 3R4, Canada

1 - Walled Garden or Cash Cow? Strategizing with Supporting Services

Bin Hu, University of Texas at Dallas, Richardson, TX, United States, Yulin Hu, Mingliu Chen

We investigate when a firm that possesses a proprietary, value-adding "supporting service" (e.g., Tesla's Supercharger network) should keep that service exclusive to its own products ("walled-garden" strategy) versus open it to rival products and monetize it directly ("cash-cow" strategy). We model a duopoly in which the service-advantaged firm and its competitor sell substitutable products over two market stages. Early-stage customers form beliefs on the service-openness decision, which then changes the firm's decision-making environment. Using a game-theoretical model and rational expectation equilibrium concept, we explicitly characterize the equilibrium prices and service-openness decisions. We show that, even with unchanged market fundamentals, the service-advantaged firm may optimally switch from exclusivity in the early stage—using the service to boost product demand—to openness in the mature stage to harvest service revenue. We then endogenize the service-advantaged firm's investment in enhancing service quality and study government mandates that force openness either immediately or with a delay. Mandates can crowd out investment when the service is only mildly lucrative, but can stimulate higher investment and raise social welfare when service revenues are substantial; the welfare-maximizing mandate (none, immediate, or delayed) varies systematically with the service charge, the intrinsic service advantage, and the cost of investment. We also show that allowing the firm to commit credibly to future openness has a negligible policy impact.

2 - Optimality of Lower-Censorship in Information Design

Yunzong Xu, University of Illinois, Champaign, IL, United States, Zhangyi Liu, Yunbei Xu

We examine the surprisingly broad optimality of "lower-censorship" rules in information design, offering a unified framework that extends and connects prior results in rating, delegation, and persuasion.

3 - Non-Exclusive Notifications: From Theory to Practice

Farbod Ekbatani, Chicago Booth School of Business, Chicago, IL, United States, Rad Niazadeh

Ride-hailing platforms face increasing uncertainty in driver compliance, with growing rates of ignored or rejected ride assignments. To address this, platforms are shifting from traditional one-to-one matching toward non-exclusive notifications, where the same ride request is sent to multiple drivers simultaneously. This paper introduces a formal framework for analyzing such notifications, capturing the trade-offs between quality and speed of the matches. With many-to-one matching, a new challenge arises: contention resolution among drivers who accept the same ride. We study two contention-resolution strategies—First Accept, which prioritizes speed by assigning the ride to the first responder, and Best Accept, which prioritizes quality by waiting to select the best available driver. We develop algorithms for selective notifications, including optimal formulations, greedy approximations, and heuristic approaches, and analyze their performance through fluid and batching models. Our results reveal a Pareto frontier between match quality and waiting time, showing when non-exclusive notification improves reliability and when it creates inefficiencies. By bridging centralized matching models with the realities of voluntary participation and driver lapses, this work provides theoretical and practical guidance for designing robust assignment protocols in ride-hailing platforms.

4 - Online Strategic Decision Making with Instruments

Rui Ai, MIT, Cambridge, MA, United States, David Simchi-Levi

This talk explores algorithmic approaches to learning and decision-making in strategic environments characterized by confounding, information asymmetry, and knowledge transfer challenges. In many economic systems, such as service fee pricing, we only observe equilibrium outcomes rather than the full demand curve, making OLS-based inference unreliable. We address this by leveraging proactive actions as instrumental variables and reveal how market randomness can aid identification, uncovering a phase transition phenomenon. When agents are forward-looking and strategic, we design a low-switching algorithm that achieves optimal regret bounds across several settings. The second part of the talk focuses on online reinforcement learning under information asymmetry, where agents act based on private signals. We examine the interplay between confounded observations and the difficulty of transferring knowledge across environments. We introduce a sample-efficient algorithm that not only identifies hidden system dynamics but also learns an ϵ -optimal policy under strategic interaction and non-i.i.d. actions. Our approach achieves near-optimal sample complexity and contributes to a broader understanding of learning in dynamic, multi-agent systems. Together, these works offer insights into how proactive experimentation and carefully designed algorithms can overcome deep structural challenges in learning and inference within strategic marketplaces.

5 - Online Mechanism Design for Crowdsourcing Systems

Renfei Tan, Massachusetts Institute of Technology, Cambridge, MA, United States

In the era of artificial intelligence, the need for high-quality labeled data has become more pressing than ever, and crowdsourcing platforms are a key source of such data. Yet these crowdsourced annotations can be vulnerable to data poisoning attacks and varying levels of worker reliability. In this talk, we introduce an online mechanism design framework that models this problem through a label-efficient and stochastic version of the prediction with expert advice paradigm. By selectively querying true labels while minimizing costs, our approach aims to maintain robust learning guarantees in the presence of adversarial or erroneous annotators. We discuss theoretical bounds, algorithmic insights, and the practical implications of applying these methods to modern crowdsourcing platforms.

TA55

Building B Level 3 B314

Analytics for Assortment, Inventory, and Last-Mile Fulfillment

Invited Session

Revenue Management

Chair: Bradley Sturt, University of Illinois Chicago, Chicago, IL, United States

1 - Approximation Schemes for Dynamic Pricing with Menu Costs**Jacob Feldman, Washington University, Clayton, MO, United States, Danny Segev**

We study a multiperiod, multiproduct, dynamic pricing problem with price adjustment costs known as “menu costs.” In this setting, decision makers, such as retailers or online platforms, incur a fixed menu cost whenever prices are updated between periods as well as a variable menu cost that linearly scales with the number of between-period price changes. Our first main contribution consists of establishing fundamental hardness results for the dynamic pricing problem of interest, unveiling inherent computational hurdles even when stripped down to its simplest form. On the positive side, we identify grocery retail as a particular application domain whose distinguishing features allow us to develop a fully polynomial time approximation scheme (FPTAS), through which optimal expected revenues are shown to be approachable within any degree of accuracy. Finally, we present two extensive numerical experiments aimed at studying the impacts of menu costs, albeit from distinct vantage points. Both studies make use of a large and diverse test bed of problem instances that we painstakingly construct from a NielsenIQ scanner grocery data set. Given the novelty of incorporating menu costs within dynamic pricing environments, we use the first set of numerical experiments to present what constitutes one of the first close looks at how these price adjustment costs shape pricing patterns and influence optimal gross margins in a multiproduct setting. The second set of experiments is used to evaluate the efficacy of our FPTAS, ultimately showing its mean revenue performance to be within 1%–2% of optimal.

2 - Assortment Optimization and the Sample Average Approximation**Bradley Sturt, University of Illinois Chicago, Chicago, IL, United States**

We study a simple approach to solving assortment optimization problems under general choice models. The approach uses samples to construct a ranking-based choice model that serves as a proxy for the true choice model, followed by finding an assortment that is optimal with respect to that proxy. In this talk, we make that approach more practically viable by developing faster algorithms for finding assortments that are optimal under ranking-based choice models. We show using real-world data that our algorithms—without any warm starts or parameter tuning—can offer more than a 30x speedup over the previous state-of-the-art. Equipped with these algorithms, we will discuss the viability of using the sample average approximation to solve assortment optimization problems in various settings of practical importance for which no polynomial time algorithms exist.

3 - Dynamic Resource Allocation with Recovering Rewards**Mika Sumida, University of Southern California, Los Angeles, CA, United States**

In many resource allocation settings, the value derived from resources depends on their usage history, with resources that have sufficient recovery or idle periods between allocations often providing greater utility or reward. This paper studies a resource allocation problem in which resource rewards recover over time following each use. Motivated by settings such as content recommendation, service platforms, and renewable energy management, we consider a dynamic matching problem with non-stationary arrivals, where customer types and matching preferences vary over time. Each arriving customer must be immediately and irrevocably matched to a resource or lost. The reward from matching a resource is non-decreasing in the time since its previous use. The goal is to maximize the expected reward collected from all arrivals over a finite time horizon.

Computing the optimal policy for this problem is computationally intractable as the number of resources grows. While the greedy policy achieves at least 1/2 of the optimal expected revenue for subadditive reward functions, it performs arbitrarily poorly for general non-decreasing reward functions. To address this, we propose an approximate policy based on a novel value function approximation that achieves a 1/3 performance guarantee for general non-decreasing reward functions. We show that for the special case of concave reward functions as well as the special case where all resources are reusable, the approximate policy's guarantee improves to 1/2. Our approximation leverages resource-specific dynamic programs that incorporate each resource's recovery level as a state, a framework that we believe to be new to the literature.

4 - Driving Efficiency: Optimizing Workforce for Last-Mile Deliveries**Tamar Cohen-Hillel, UBC Sauder School of Business, Vancouver, BC, Canada, Liron Yedidsion, Tolga Cezik**

Last-mile delivery labor planning is a critical challenge for logistics providers like Amazon, where timely delivery and cost efficiency are essential. In this work, we address the problem of determining the optimal number of drivers required each week over a given planning horizon, accounting for uncertain demand and cross-time constraints. We formulate this as a stochastic dynamic program, capturing both fixed hiring costs and variable operational costs.

To address the computational challenges of this NP-hard problem, we propose a dynamic programming-based approximation algorithm. Our approach effectively balances the trade-off between cost efficiency and service reliability, minimizing the total expected cost while ensuring all shipments are delivered on time. The cost model accounts for the fixed expenses of onboarding new drivers and the operational costs of dedicated vans and flex drivers, providing a realistic representation of the cost structure.

We validate our approach with both theoretical bounds and extensive computational experiments, demonstrating that our algorithm achieves a tight optimality gap and performs near-optimally in practice. These results highlight the potential for significant cost savings in last-mile delivery operations through intelligent labor planning, offering valuable insights for both practitioners and researchers.

TA56

Building B Level 3 B315

Optimization and Strategy Issue in Matching Markets

Invited Session

Auctions and Market Design

Chair: Chengyue He, Columbia University, Jersey City, NJ, 10027, United States

1 - Minimum Cut Representability of Stable Matching Problems**Chengyue He, Columbia University, New York, NY, United States**

We introduce and study Minimum Cut Representability, a framework to solve optimization and feasibility problems over stable matchings by representing them as minimum s-t cut problems on digraphs over rotations. We provide necessary and sufficient conditions on objective functions and feasibility sets for problems to be minimum cut representable. In particular, we define the concepts of first and second order differentials of a function over stable matchings and show that a problem is minimum cut representable if and only if, roughly speaking, the objective function can be expressed solely using these differentials, and the feasibility set is a sublattice of the stable matching lattice. To demonstrate the practical relevance of our framework, we study a range of real-world applications, including problems involving school choice with siblings and a two-stage stochastic stable matching problem. We show how our framework can be used to help solving these problems.

2 - Adding Private Information into Centralized Algorithms: A Field Experiment at A Ride-Sharing Platform**Yingzhi Liang, Naveen Jindal School of Management, University of Texas at Dallas, Dallas, TX, United States**

Many have seen the gig economy as the “future of work”. Despite having a boss-free working environment and a flexible working schedule, gig workers have little power over the algorithm that assigns them tasks. Take ride-sharing drivers for example, the platform assigns drivers trips without considering driver preferences for locations. As a result, drivers might suffer from a lack of working location flexibility. Through a field experiment on the largest ride-sharing platform in China, I examine the effect

of incorporating location preferences into the centralized matching. I find that allowing drivers to self-define working regions increases their working hours and income by more than 4%, without lowering their productivity. Drivers use this feature both for necessity constraints and strategy concerns.

3 - Managing Self-Interested Behavior in Failure-Aware Kidney Exchange

Itai Feigenbaum, City University of New York, New York, NY, United States

Patients requiring kidney transplant may have proxy donors, who wish to donate a kidney to the patient- but are incompatible. However, patients can swap their proxy donors for a compatible one. Transplants often get canceled, so planned swaps should maximize the expected number of non-canceled transplants. Exact maximization is neither strategyproof nor core-stable: patients can increase their probability of receiving a kidney via hiding proxy donors, or via side-arrangements. We introduce a group-strategyproof and core-stable mechanism called SuperGreedy. On uniform markets, where cancellation probabilities are equal across transplants, SuperGreedy is worst-case 2.03-approximate for realistically large cancellation probabilities, and empirically average-case 1.142-approximate in simulation. Our results extend beyond current practice, to when patients can swap multiple proxy donors, with weaker theoretical approximation guarantees but even stronger empirical performance. SuperGreedy is efficiently computable when restricted to size-2 cycles; without this restriction SuperGreedy is NP-hard, but still practically computable on uniform markets.

TA57

Building B Level 3 B316

Learning in Auctions and Games

Invited Session

Auctions and Market Design

Chair: Gabriele Farina, MIT, Cambridge, MA, United States

Co-Chair: Giannis Fikioris, Cornell University, Ithaca, NY, United States

1 - Learning in Budgeted Auctions with Spacing Objectives

Giannis Fikioris, Cornell University, Ithaca, NY, United States

In many repeated auction settings, participants care not only about how frequently they win but also how their winnings are distributed over time. This problem arises in various practical domains where avoiding congested demand is crucial, such as advertising campaigns that require sustained visibility over time. We introduce a simple model of this phenomenon, modeling it as a budgeted auction where the value of a win is a concave function of the time since the last win. This implies that for a given number of wins, even spacing over time is optimal. We also extend our model and results to the case when not all wins result in "conversions" (realization of actual gains), and the probability of conversion depends on a context. The goal is to maximize and evenly space conversions rather than just wins.

We study the optimal policies for this setting in second-price auctions and offer learning algorithms for the bidders that achieve low regret against the optimal bidding policy in a Bayesian online setting. Our main result is a computationally efficient online learning algorithm that achieves $O(\sqrt{T})$ regret. We achieve this by showing that an infinite-horizon Markov decision process (MDP) with the budget constraint in expectation is essentially equivalent to our problem, even when limiting that MDP to a very small number of states. The algorithm achieves low regret by learning a bidding policy that chooses bids as a function of the context and the system's state, which will be the time elapsed since the last conversion.

2 - Robust Online Fair Allocation with a Single Sample

Zongjun Yang, Columbia University, New York, NY, United States, Christian Kroer, Rachitesh Kumar

We consider the online fair allocation problem with T sequentially arriving items and n agents with linear, additive utilities. Given potentially non-stationary or adversarial arrivals over time, prior works either focuses on worst-case approximation ratio results, or assumes a bounded extent of non-stationarity. We investigate the question of using a minimum number of historical sample traces to enable no-regret guarantees for the Eisenberg-Gale objective, with arbitrarily non-stationary input. We show that a single sample per distribution is sufficient for the resolving algorithm to achieve $O(\log T/\sqrt{T})$ regret, while being robust to potential errors in the sampling distribution.

3 - Equilibria in Uniform Price Auctions: Theory, Computation, and Market Insights

Qinzheng Li, National University of Singapore, Singapore, Singapore, Yifan Feng, Negin Golrezaei

We study pure-strategy Nash equilibrium outcomes of multi-unit auctions under the $\$K\$$ -th price payment rule. Despite their widespread use in resource allocation markets such as treasury bond sales, electricity markets, and carbon credit trading, these auctions remain poorly understood due to their combinatorial complexity and bidders' strategic incentives for bid shading. In this paper, we provide a complete characterization of the set of allocation and clearing price outcomes at equilibrium. This enables us to develop a mixed-integer programming formulation to compute Nash equilibrium outcomes and select those with desirable properties, such as maximum welfare or revenue. We further extend our analysis to incorporate reserve prices. We then leverage our characterization to study market stability through numerical experiments, which show that markets with fewer bidders and coarser bid grids are more prone to tacit collusion. These inefficiencies can be mitigated by increasing the number of bidders to enhance competition or by using reserve prices to curb revenue loss. We also examine repeated multi-unit auctions and find that adaptive bidding algorithms often converge to stable equilibria and exhibit inertia when initialized at an equilibrium—suggesting that initial bids can serve as a practical lever to avoid undesirable outcomes.

4 - Learning to Play Against Unknown Opponents

Jon Schneider, Google, Columbia, NY, United States

We consider the problem of a learning agent who has to repeatedly play a general sum game against a strategic opponent who acts to maximize their own payoff by optimally responding against the learner's algorithm. The learning agent knows their own payoff function, but is uncertain about the payoff of their opponent (knowing only that it is drawn from some distribution D). What learning algorithm should the agent run in order to maximize their own total utility, either in expectation or in the worst-case over D ? When the learning algorithm is constrained to be a no-regret algorithm, we demonstrate how to efficiently construct an optimal learning algorithm (asymptotically achieving the optimal utility) in polynomial time for both the in-expectation and worst-case problems, independent of any other assumptions. When the learning algorithm is not constrained to no-regret, we show how to construct an ϵ -optimal learning algorithm (obtaining average utility within ϵ of the optimal utility) for both the in-expectation and worst-case problems in time polynomial in the size of the input and $1/\epsilon$, when either the size of the game or the support of D is constant. Finally, where the learner wishes to maximize their minimum payoff over all possible optimizer types, we construct a learner algorithm that runs in polynomial time in each step and guarantees convergence to the optimal learner payoff. All of these results make use of recently developed machinery that converts the analysis of learning algorithms to the study of the class of corresponding geometric objects known as menus.

TA58

Building B Level 3 B317

Smart Supply Chain Systems

Contributed Session

Contributed

Chair: Xiao Xiao, California State University, Stanislaus, Turlock, CA, United States

1 - Spillover Effect of Supply Chain Credit Risk :Empirical Evidence from Chinese a-Share Listed Enterprises

Yu Liu, Shanghai Jiaotong University, Shanghai, China, People's Republic of

Based on the data of Chinese A-share listed enterprises from 2013 to 2022, this paper selected core enterprises and small and medium-sized enterprises (SMEs) with supply relationship as the research samples. calculated the credit risk level by using the KMV model, researched the spillover effect of credit risk from core enterprises to SMEs in the supply chain, explored the transmission path of supply chain credit risk, and analyzed the heterogeneity of the spillover effect of credit risk in the supply chain. The research results show that there is a spillover effect of credit risk from core enterprises to associated SMEs in the supply chain. Further, the good credit performance of core enterprises can significantly improve the profitability, operating ability, and debt-servicing ability of associated SMEs, and thus mitigate the spillover effect of supply chain credit risk; The spillover effect of credit risk in supply chain is more obvious under high financing constraints, and high supply chain concentration exacerbates the spillover effect of credit risk in supply chain. The findings offers empirical evidence for related enterprises to develop prevention and control strategies for supply chain credit risk and provides practical insights for government agencies to strengthen credit supervision and improve the financing environment.

2 - Tracing the Paths of Cascading Disruptions in Multi-Tier Supply Networks

Phu Nguyen, Berlin School of Economics and Law, Berlin, Germany, Dmitry Ivanov

Modern supply chains have grown increasingly complex and fragile, exposing firms to cascading disruptions, also known as ripple effects. While prior research has focused on identifying critical suppliers and materials to enhance resilience, little research has been done to understand the structure and characteristics of disruption propagation paths. Our study addresses this gap by proposing an approach to analyze the critical disruption paths in a deep-tier supply network. We model the supply chain as a random network and employ a simulation framework based on the Susceptible–Infected–Recovered paradigm to capture the supply chain emergence over time. To systematically examine the properties of critical disruption paths, we conduct a structured experiment design using three distinct disruption scenarios, each representing different levels of impact severity. Our approach is further validated through an empirical case study involving a leading European manufacturing firm. The supply network comprises a focal firm, 130 customers, 251 direct suppliers, and 1,479 Tier-2 suppliers. The results reveal key structural attributes of critical disruption paths significantly influencing the focal firm's exposure. Our findings inform the development of a minimum viable supply chain stress-testing model in research. Supply chain managers can leverage the results to safeguard critical paths and ensure more efficient allocations of resilience resources.

3 - Mitigating Supply Chain Risks Under Multi-Level Geopolitical Tensions: the Impact of Digital Network Capability on Supply Chain Resilience

AI-Hsuan Chiang, Ming Chuan University, Taipei, Taiwan

Amid rising geopolitical tensions, natural disasters, and regional conflicts, global market uncertainty continues to grow, significantly heightening the risk of supply chain disruptions. Under multilayered geopolitical risks, firms encounter complex challenges at multiple levels. Firm-level risks, such as regulatory interference or trade restrictions, directly affect operations. In contrast, regional or national-level risks, including cross-border sanctions or abrupt policy shifts (e.g., trade wars), disrupt broader supply chain environments. To enhance responsiveness and resilience, firms are increasingly adopting advanced technologies such as Industry 4.0 and the Internet of Things (IoT). These innovations are enabling the transformation of traditional supply chains into digital ones, thereby strengthening digital network capabilities to absorb external shocks.

Grounded in the dynamic capabilities perspective, this study investigates how digital network capability contributes to supply chain resilience, helping firms mitigate the effects of multilayered geopolitical risks. The study also identifies key antecedents that shape digital network capability. Data were collected from firms across multiple countries and analyzed using structural equation modeling. The results indicate that digital network capability significantly enhances supply chain resilience, which in turn improves both operational and environmental performance. Moreover, resilience acts as a key mediator between digital network capability and performance outcomes. Notably, while resilience buffers firm-level geopolitical risks, it remains susceptible to erosion under regional-level geopolitical pressure. These findings underscore the dual role of digital network capability—serving both as a product of strategic enablers and as a foundation for organizational resilience in volatile geopolitical contexts.

4 - Who Rules the Battery? Technology Control, Supply Chain Synergy, and Cost Structure Dynamics in Vehicle-Battery Separation Models

Bo Liu, Southeast University, Nanjing, China, People's Republic of, Yi Ding

The vehicle-battery separation (VBS) model represents a crucial innovation in the new energy vehicle industry, where its supply chain structure and operational strategies systematically influence battery leasing prices, technical performance, and corporate profits. This study conducts a comparative analysis of three business models: Manufacturer-Integration (MI), Supplier-Integration (SI), and Manufacturer-Cooperation (MC), revealing the dynamic equilibrium between synergistic effects, technological control, and economies of scale across different operational approaches. The research demonstrates that in terms of pricing mechanisms, the SI model establishes distinct leasing price advantages by capitalizing on suppliers' specialized expertise and automakers' cross-subsidization strategies. Regarding technical performance, the MI strategy delivers optimal results in high-efficiency environments through fully integrated industrial chains, while the MC strategy shows greater adaptability in low-efficiency or high revenue-sharing ratio scenarios. Profit distribution exhibits threshold effects related to technological control and economies of scale: MI generates maximum profits in high-efficiency phases, SI prevails in low-efficiency stages, and MC proves most competitive during moderate efficiency periods or when vehicle body revenues are substantial. The study also identifies how consumer risk preferences asymmetrically influence market equilibrium: under low R&D costs, risk aversion paradoxically elevates market prices and profit margins, whereas high-cost conditions may trigger market failures characterized by "performance degradation-price escalation". This research establishes a strategic decision framework for VBS model adoption: technology-leading enterprises should implement MI models for value capture; while cost-sensitive operators should optimize efficiency via either SI or MC approaches.

5 - A Multi-Agent Based Digital Twin Framework for the Digitalization of Assembly Environments

Michail Katsigiannis, Auburn University, Auburn, AL, United States, Konstantinos Mykoniatis

This work introduces a scalable Multi-Agent-Based Digital Twin (MABDT) framework designed to create Digital Twin models that can monitor complex production systems. The framework architecture is structured into three interoperable layers: the Physical Layer, the Digital Twin Layer, and the Application Layer. The Physical Layer captures live data from sensors, automation systems, and Internet of Things devices. Next, the Digital Twin Layer comprises autonomous agents that maintain state charts, process incoming messages, and simulate future scenarios. Finally, the Application Layer delivers analytics, visualization, and decision-support tools. Communication Agents orchestrate bi-directional data flows, translating raw inputs into structured events for Digital Twin Agents, which can issue control commands back to the physical environment. Digital Twin Agents leverage state-chart-based models to represent entities, such as workstations, products, and handling resources, defining clear state transitions, communication interfaces, and action routines. The proposed MABDT framework advances the Digital Twin paradigm toward highly adaptive, autonomous, and interoperable simulation ecosystems by focusing on modular agent definitions and flexible communication patterns. A case study at the Tiger Motors educational assembly line at Auburn University illustrates the framework's use in modeling a 15-workstation mixed-model assembly line. Car, Workstation, and Inspection Station Agents provide real-time tracking of vehicle progression, workstation utilization, and fault events, delivering dynamic performance metrics and interactive visualizations for teaching and prototyping digital manufacturing processes.

6 - Risk Sharing in a Two-Level Supply Chain with Variable Capacity and Random Yield

Xiao Xiao, California State University, Stanislaus, Turlock, CA, United States

Research on production uncertainty is crucial in today's complex and interconnected business environment, as supply chains contend with numerous uncertainties, including demand fluctuations and supplier disruptions, significantly affecting their operations. This study is based on three centralized supply chains and three decentralized supply chains, and investigates two types of production uncertainty: Variable Capacity (VC) and Random Yield (RY), examining their impacts on supply capability, relationships among supply chain members, decision-making behaviors, and overall performance. RY, often resulting from imperfect processes, is predictable and measurable, unlike VC, which is triggered by entirely random factors such as unforeseen interruptions and unplanned maintenance, rendering it uncontrollable. Our findings reveal that centralized models are predominantly influenced by RY due to its predictability, enabling manufacturers to adjust their production strategies accordingly. In contrast, the unpredictable nature of VC, stemming from external shocks, is frequently disregarded in decision-making. In the absence of risk-sharing, RY primarily affects manufacturers, with VC having a minimal impact. However, RY influences manufacturers' production decisions, while retailers' ordering strategies consider both types of uncertainty. In response to production uncertainties, manufacturers adapt to RY's predictability within supply chains, whereas retailers modify their ordering practices to mitigate risks arising from both supply and demand uncertainties. These insights underscore the distinct impacts of production uncertainties on supply chain entities and highlight management practices for addressing these challenges.

TA59

Building B Level 3 B318

Learning and Optimization in Healthcare Operations

Contributed Session

Contributed

Chair: Marco Patino, North Dakota State University, Fargo, ND, United States

1 - Multi-Item Inventory Optimization with Correlated Demands

Zhaleh Rahimi, McMaster University, Hamilton, ON, Canada, Douglas Down, Na Li

We study optimal ordering policies for a multi-item periodic-review inventory system, considering demand correlations and historical data for the products involved. While previous studies have considered demand correlation, the effects of different correlation types, such as autocorrelation, cross-correlation and time-lagged dependencies, is underexplored. To address this, we model the demand process using a VARMA (vector autoregressive moving average) framework. VARMA can effectively capture these correlations and enable the exact computation of optimal policies under separable cost functions. We also investigate whether a VARMA process can be effectively fitted to operational data through numerical experiments using both synthetic data and real-world demand for immunoglobulin blood products including intravenous (IVIg) and subcutaneous (SCIg).

2 - Multi-Objective Evolutionary Optimization for Healthcare Districting: a Customized Nsga-Ii Approach

Huijae Kim, KAIST, Daejeon, Korea, Republic of, Taesik Lee

Properly districting a large service region—such as a country—into sub-regions is essential for the efficient and equitable delivery of healthcare services. Healthcare districting inherently involves trade-offs between two critical objectives: *self-containment* (i.e., the proportion of patients treated within their own district) and *accessibility*. This study proposes a bi-objective optimization model that aims to maximize the average self-containment rate while minimizing the maximum intra-district travel distance, under constraints on population size, healthcare facility capacity, and geographic contiguity. To solve this model, we design a tailored NSGA-II algorithm that incorporates an efficient encoding structure inspired by graph-partitioning problem and domain-specific genetic operators to enhance both search efficiency and solution quality. Using Korean National Health Insurance claims data, we conduct extensive experiments. For small-scale instances, we validate our approach against exact optimization methods to ensure solution accuracy. For larger-scale problems, we benchmark our method against two established NSGA-II variants, demonstrating the superior quality of Pareto front and faster convergence. The results highlight the practical utility of multi-objective evolutionary optimization for addressing real-world healthcare districting problems.

3 - Early Breast Cancer Detection in Rural North Dakota: Strategic Resource Allocation

Marco Patino, North Dakota State University, Fargo, ND, United States, Diana Lopez-Soto, Ph.D.

Cancer remains one of the most significant public health concerns in the United States, with approximately half of the men and one-third of women expected to develop the disease in their lifetime. Among all types, breast cancer in particular stands out due to its high prevalence in new cases and cost. In the U.S., over 300,000 new cases are projected in 2025, with treatment expenses reaching nearly 30 billion in 2020. While early detection can significantly improve survival rates and reduce costs by up to one-third, many individuals face barriers to timely screening.

Out-of-pocket expenses and limited access to healthcare services contribute to delayed diagnoses. In some cases, treatment costs for breast cancer can exceed \$120,000. These challenges are especially pronounced in North Dakota, where breast cancer incidence rates among women are above the national average, and rural populations often lack access to preventive care.

This study proposes the use of mathematical programming to strategically deploy mobile health units across North Dakota. The goal is to improve early detection rates and reduce both mortality and the economic burden on underserved populations. The proposed model could serve as a blueprint for other rural states facing similar disparities in cancer outcomes.

TA60

Building B Level 3 B319

LLM Applications in Decision Support

Contributed Session

Contributed

Chair: Behrooz Davazdahemami, University of Wisconsin-Whitewater, Whitewater, WI, United States

1 - Which LLM Builds the Best Optimization Model?

Brian Schaefer, SimpleRose, St. Louis, MO, United States

This talk compares leading large language models (LLMs)—including ChatGPT, Claude, and Gemini—for their ability to formulate, debug, and explain optimization models. Through a series of structured prompts and real-world modeling tasks, I'll highlight where each model excels or struggles, from translating problem statements to generating clean code or catching feasibility issues. The session will include practical tips for better prompting and model verification, and end with guidance on which LLM to use for which optimization task.

2 - Automatic Validation of Optimization Models Using LLMs

Alexander Zadorojniy, IBM Research, Haifa, Israel, Segev Wasserkrug, Eitan Farchi

Recently, using Large Language Models (LLMs) to generate optimization models has become increasingly popular. However, despite the potential benefits of such generation, a major open question is how to validate the correctness of the generated models. In our work, we propose a novel method for automatic validation of such LLM generated optimization models. Our approach builds upon best practices in software testing and using the natural language description of the problem to be modelled, consists of several LLM based agents: an agent that uses an LLM to create a problem level, rather than an optimization model level, testing interface; an LLM agent to create unit tests; and an LLM agent to generate mutations with the goal of helping the unit test creator cover corner cases based on the problem description. In our context mutations are applied to the constraints of a given problem in the benchmark. For example, a constraint of a given problem in the benchmark may state that $x+y \leq 8$ and its mutant may be $x+y \leq 7$. We show empirically that our approach results in higher coverage of the generated optimization models with respect to standard methods such as objective function value comparison.

3 - AI Decision Support: Alignment of Risk Attitudes

Elisabeth Paté-Cornell, Stanford University, Stanford, CA, United States

AI decision support includes both the relevant information and a suggestion (or implementation) of an optimal decision option. Information can be checked against facts and beliefs. The choice of an option, however, has to be based on relevant preferences. This talk will discuss the need and the implementation in AI decisions of a risk attitude that fits that of the decision maker, and will present model of examples to do so.

4 - E2r2: a Hybrid Explainable Framework for Case-Based Reasoning and Generative Decision Support

Behrooz Davazdahemami, University of Wisconsin-Whitewater, Whitewater, WI, United States, Hamed Zolbanin, Dursun Delen

The rapid evolution of artificial intelligence has introduced new methodological possibilities for Information Systems Research (ISR), particularly in decision support. This paper presents E2R2—an Explain, Embed, Retrieve, and Reason framework—that integrates explainable AI, semantic similarity matching, and generative reasoning to support interpretable, example-driven classification. E2R2 was applied to the task of predicting second-year college attrition using a real-world student dataset and demonstrated strong predictive accuracy (84%), session-level consistency (92%), and robustness under partial data conditions (88% agreement).

The framework operates in three stages: (1) model prediction and SHAP-based feature attribution, (2) generation of natural language rationales via a fine-tuned generative model (o3-mini), and (3) classification and explanation of new instances based on semantic proximity to precedent cases. Unlike traditional black-box classifiers or basic CBR systems, E2R2 provides personalized, SHAP-informed explanations grounded in contextual case analogies, enabling transparent, cognitively aligned decision support.

Experimental results confirm that E2R2 not only performs reliably across different execution contexts but also delivers actionable, domain-relevant insights in natural language. The system balances computational precision with interpretive relevance—offering a pragmatic approach to the “precision vs. relevance” tradeoff outlined by Zadeh. The study advances ISR by operationalizing a flexible and reusable methodology suitable for digital coaching, algorithmic decision support, and adaptive information systems. It also lays the groundwork for future enhancements, including user-in-the-loop refinement and domain adaptation.

This work contributes to the design and evaluation of AI-driven decision methodologies that are interpretable, resilient, and grounded in human-centered reasoning.

TA61

Building B Level 4 B401

Recent Advances in High-Dimensional Change Detection and Gaussian Processes

Invited Session

Quality, Statistics and Reliability

Chair: Ruizhi Zhang, University of Georgia, Athens, GA, 30542, United States

Co-Chair: Liyan Xie, University of Minnesota, 1400 Park Avenue, MINNEAPOLIS, 55404, United States

1 - High-Dimensional Dynamic Process Monitoring by Pca-Based Sequential Learning

Xiulin Xie, Florida State University, Tallahassee, FL, United States

Air pollution surveillance is critically important for public health. To monitor a sequential process online, a major statistical tool is statistical process control (SPC) chart. However, traditional SPC charts are developed mainly for monitoring production lines in the manufacturing industry under the assumptions that process observations at different observation times are independent and identically distributed with a parametric (e.g., normal) distribution when the process is stable. Nevertheless, the air pollution and meteorological data would not satisfy these conditions due to serial correlation, seasonality, and other complex data structure. In this talk, we present our latest research on sequential monitoring of air quality over time. In particular, we propose a flexible method for sequential monitoring of high-dimensional dynamic processes with serially correlated data. The new method is based on nonparametric longitudinal modeling for describing the longitudinal pattern of the process under monitoring, principal component analysis for dimension reduction, and a sequential learning algorithm for developing an effective decision rule. Numerical studies and real data applications show that the proposed method works well.

2 - Adaptive Block-Based Change-Point Detection for Sparse Spatially Clustered Data with Applications in Remote Sensing Imaging

Lynna Chu, Iowa State University, Ames, IA, United States, Alan Moore, Zhengyuan Zhu

We present a non-parametric change-point detection approach to detect potentially sparse changes in a time series of high-dimensional observations or non-Euclidean data objects. We target a change in distribution that occurs in a small, unknown subset of dimensions, where these dimensions may be correlated. Our work is motivated by a remote sensing application, where changes occur in small, spatially clustered regions over time. An adaptive block-based change-point detection framework is proposed that accounts for spatial dependencies across dimensions and leverages these dependencies to boost detection power and improve estimation accuracy. Through simulation studies, we demonstrate that our approach has superior performance in detecting sparse changes in datasets with spatial or local group structures. An application of the proposed method detect activity, such as new construction, in remote sensing imagery of the Natanz Nuclear facility in Iran is explored.

3 - Change Point Detection for Multiple Dynamic Processes with Exogenous Covariates

Zibo Tian, University of Florida, Gainesville, FL, United States, Peihua Qiu

Detecting structural changes in dynamic processes is important in fields like finance, genomics, and manufacturing, where abrupt changes in system dynamics can signal regime changes, faults, or external influences. Traditional change point detection methods focus on identifying transitions in univariate or multivariate processes but often overlook the role of exogenous covariates, which can drive underlying structural changes. To address this gap, this paper develops a change point detection method for piecewise stationary vector autoregressive models with exogenous covariates (VARX) to address structural changes in the observed multivariate time series influenced by exogenous variables. Unlike existing methods focusing on single-process settings, our approach can detect change points shared across multiple processes with heterogeneous exogenous effects. By reformulating change point detection as a variable selection problem, a sparse group Lasso-based approach is employed in our method to identify candidate change points, refined subsequently by a backward elimination procedure. This enables simultaneous estimation of the number, locations, and other parameters of the structural changes in high-dimensional settings. Numerical studies show that the proposed method performs well in various scenarios considered.

4 - Sequential Change Point Detection via Denoising Score Matching

Liyan Xie, University of Minnesota, MINNEAPOLIS, MN, United States

Sequential change-point detection plays a critical role in numerous real-world applications, where timely identification of distributional shifts can greatly mitigate adverse outcomes. Classical methods commonly rely on parametric density assumptions of pre- and post-change distributions, limiting their effectiveness for high-dimensional, complex data streams. This paper proposes a score-based CUSUM change-point detection, in which the score functions of the data distribution are estimated by injecting noise and applying denoising score matching. We consider both offline and online versions of score estimation. Through theoretical analysis, we demonstrate that denoising score matching can enhance detection power by effectively controlling the injected noise scale. Finally, we validate the practical efficacy of our method through numerical experiments on two synthetic datasets and a real-world earthquake precursor detection task, demonstrating its effectiveness in real scenarios.

5 - Scalable Gaussian Processes with Low-Rank Deep Kernel Decomposition

Yunqin Zhu, Georgia Institute of Technology, Atlanta, GA, United States, Henry Yuchi, Yao Xie

Kernels are key to encoding prior beliefs and data structures in Gaussian process (GP) models. The design of expressive and scalable kernels has garnered significant research attention. Deep kernel learning enhances kernel flexibility by feeding inputs through a neural network before applying a standard parametric form. However, this approach remains limited by the choice of base kernels, inherits high inference costs, and often demands sparse approximations. Drawing on Mercer's theorem, we introduce a fully data-driven, scalable deep kernel representation where a neural network directly represents a low-rank kernel through a small set of basis functions. This construction enables highly efficient exact GP inference in linear time and memory without invoking inducing points. It also supports scalable mini-batch training based on a principled variational inference framework. We further propose a simple variance correction procedure to guard against overconfidence in uncertainty estimates. Experiments on synthetic and real-world data demonstrate the advantages of our deep kernel GP in terms of predictive accuracy, uncertainty quantification, and computational efficiency.

TA62

Building B Level 4 B402

Recent Advances in Experimental Design and Variable Selection Techniques

Invited Session

Quality, Statistics and Reliability

Chair: William Fisher, JMP Statistical Discovery LLC, Raleigh, NC, United States

Co-Chair: Qiong Zhang, Clemson University, Clemson, SC, United States

1 - Combinatorial Testing of Engineered Systems with Hard-to-Change Factors**William Fisher, JMP, Cary, NC, United States**

The testing of engineered systems is a complex task which often requires a test engineer to select a suite of test cases to investigate in order to detect and correct errors. Test engineers often must strategically select their suite of test cases due to budgets on time and other operational constraints. When the test cases are defined by a set of parameters or factors, one method which is often used to guide the selection of a test suite is combinatorial testing, and the construct which is typically used to facilitate this is known as a covering array. In some instances of a combinatorial testing problem, there may be a subset of factors that are expensive or hard to change. In design of experiments, split-plot designs are used when hard-to-change factors are present. Borrowing the concept of split-plot designs from design of experiments, we extend the methodology of combinatorial testing by showing a method for constructing covering arrays with split-plot structure. We then demonstrate the efficiency of our method in a case study involving the testing of an implementation of XGBoost.

2 - Robust Models for Experiments with Quantitative and Qualitative Responses**Yu Wang, University of Massachusetts Amherst, Amherst, MA, United States, Lulu Kang**

In supervised learning, datasets often contain both qualitative and quantitative response variables. Existing literature demonstrates that jointly modeling these mixed-type responses leads to superior prediction accuracy compared to separate modeling approaches. However, challenges such as high-dimensionality and outliers further complicate the analysis. In this talk, we introduce a novel framework for robust regression and classification that addresses these issues. Our approach formulates a loss function based on density power divergence, ensuring robustness against outliers and model misspecification. By integrating this divergence-based loss with L1 regularization, our method enables accurate prediction of both qualitative and quantitative responses in high-dimensional settings. To solve the resulting nonconvex optimization problem, we employ a proximal gradient algorithm enhanced by the Barzilai-Borwein spectral stepsize. Finally, we validate the proposed framework through numerical simulations and real-world examples.

3 - Optimal Sparse Projection Design for Experiments with Treatment Cardinality Constraints**Kexin Xie, Virginia Tech, Blacksburg, VA, United States, Ryan Lekivetz, Xinwei Deng**

Modern experimental designs often face the so-called treatment cardinality constraint, which is the constraint on the number of included factors in each treatment. Experiments with such constraints are commonly encountered in engineering simulation, AI system tuning, and large-scale system verification. This calls for the development of adequate designs to enable statistical efficiency for modeling and analysis within feasible constraints. In this work, we propose an optimal sparse projection (OSP) design for systems with treatment cardinality constraints. We introduce a tailored optimal projection (TOP) criterion that ensures a good space-filling property in subspaces and promotes orthogonality or near-orthogonality among factors. To construct the proposed OSP design, we develop an efficient two-stage Allocation-Assignment Algorithm (AAA) based on genetic algorithm (GA) to efficiently explore the design space with treatment cardinality constraints. Numerical examples demonstrate the merits of the proposed method.

4 - Screening Designs for Continuous and Categorical Factors**Ryan Lekivetz, JMP Statistical Discovery LLC, Cary, NC, United States**

In this talk, we discuss a new class of saturated designs containing m three-level continuous factors and $m-1$ two-level categorical or continuous factors in $n=2m$ runs, where $m \geq 4$. A key advantage is that these designs are available for any even $n \geq 8$. With effect sparsity or by not making use of all of the two-level columns of the design, we demonstrate via simulation that it is possible to identify up to three active quadratic effects. When n is a multiple of 8, the designs are orthogonal. When n is a multiple of four and not a multiple of 8, the three-level factors are orthogonal to each other and to the two-level factors, and the two-level factors are nearly orthogonal to each other. Finally, when n is a multiple of two, and not a multiple of four or 8, the three-level and two-level factors are nearly orthogonal within those groupings, and orthogonal to each other.

5 - Bayesian Variable Selection for Gaussian Process Models in Computer Experiments: Joint Credible Region and Indicator-Based Approaches**fan zhang, Clemson University, Clemson, SC, United States**

Gaussian process (GP) models are commonly used in the analysis of computer experiments. Variable selection in GP models is of significant scientific interest, but existing solutions remain unsatisfactory. For each variable in a GP model, there are two potential effects with different implications: one is on the mean function, and the other is on the covariance function. However, most existing research on variable selection for GP models has focused only on one of these effects. In this talk, we introduce two Bayesian approaches for variable selection in GP models, each addressing the challenge of considering both the mean and covariance functions.

Joint credible region method for variable selection in universal kriging introduces an inert variable and constructs a joint credible region based on the posterior samples of both regression coefficients and correlation parameters of the inert variable. The method determines variable activity by examining the location of each variable's posterior samples relative to the joint credible region. Indicator-based Bayesian variable selection procedure for GP models introduces an indicator variable to simultaneously consider variable effects in the mean and covariance functions. A variable is defined to be inactive if both effects are not significant, and an indicator is used to represent whether the variable is active or not. The proposed methods provide flexible and effective tools for variable selection in GP models for computer experiments. Simulation studies and real-data applications confirm the accuracy and robustness of both approaches.

TA63

Building B Level 4 B403

Advances in Understanding AI

Contributed Session

Contributed

Chair: Kaige Gao, Cleveland, OH, United States

1 - Unveiling Shadow AI use in the Workplace: Psychological Ownership as a Driver of Employee Knowledge Hoarding and the Moderating Role of Job Autonomy**Xiao Han, Xi'an Jiaotong University, Xi' an, China, People's Republic of**

With the rapid spread of generative AI tools and the distinct increase in their ease of use, employees can easily access AI tools and use them to support their work. At the same time, many organizations have yet to develop specific guidelines or management processes regarding the use of AI tools, resulting in a lack of organizational-level control and surveillance of the work process using AI tools. In this context, shadow AI use refers to the behavior of employees using AI tools without the approval of the technology department of the organization. Existing research around shadow AI use is confined to the perspective of viewing AI as an information technology, mainly defining shadow AI use as an unethical behavior of employees, focusing on its impact on organizational knowledge leakage, and ignoring the impact of shadow AI use on employees' knowledge behavior at the individual level. From the perspective of psychological ownership theory, this study argues that employees' AI shadow use causes employees' knowledge hoarding behavior by increasing knowledge-based psychological ownership. Based on the above mediating mechanism, we further explored the positive moderating role of a specific job characteristic-job autonomy on this process from the perspective of self-determination theory. The results of a multi-stage questionnaire study from employees in China confirm the proposed research hypotheses and provide useful insights into how organizations can maintain employees' knowledge-based initiatives in the context of the increasing prevalence of AI tools at the individual employee level.

2 - Prior Knowledge Does Help: Generating Adversarial Samples with Prior-Guided Multi-Objective Optimization

Yilun Jin, Southeast University, Nanjing, China, People's Republic of

Generating adversarial text samples in the black-box and hard label setup is a challenging task, as the data space is inherently discrete and non-differentiable, and only the predicted label is accessible. Research on this problem is still in the embryonic stage and only a few methods are available. Meanwhile, current methods often encounter difficulties in simultaneously achieving low perturbation rate and high semantic similarity, and they often require a large number of queries.

To address these limitations, we propose a novel prior-guided attack (PGAttack) to generate adversarial text samples satisfying multi-objectives in a query-efficient manner. Specifically, PGAttack includes a new word importance-aware method to identify critical words within the text. Then, a new multi-objective optimization algorithm is proposed to generate adversarial text based on this prior knowledge. This algorithm draws inspiration from evolutionary computation and incorporates two customized prior-guided operators and a new selection criterion. Extensive experimental results on four classical text classification datasets and across three widely-used text models, demonstrate the superiority of PGAttack. Furthermore, evaluations on two real-world APIs demonstrate that PGAttack can generate better adversarial samples while reducing query costs by more than 50% compared to baseline attack methods.

3 - Does AI Complement R and D Investments in Driving Innovation?

Patricia Angle, University of Oklahoma, Norman, OK, United States, Lubomir Litov

In 2024, U.S. private equity and venture capital investment in generative AI soared to \$56 billion, nearly doubling investments of the previous year, with organizations reporting significant revenue gains associated with generative AI use (S&P Global, 2025). However, the rapid integration of AI raises challenges, particularly for R&D-intensive firms, as AI's opaque algorithms complicate decision-making transparency. This paper investigates the impact of AI adoption on firm innovation, focusing on the role of R&D investment as a moderating factor. Utilizing data from the Aberdeen Ci Technology Database (2013–2017) and the USPTO (2013–2020), the study examines the relationship between AI-related software adoption and firm-level innovation outcomes, measured through patent grants. Fixed-effects Poisson panel regressions reveal a significant positive relationship between business intelligence (BI) intensity and patents, especially with a two- to three-year lag. However, the effect diminishes after winsorization, indicating that the positive relationship is driven by a small subset of firms with high BI intensity. Further analysis highlights robust complementarity between BI intensity and R&D intensity, suggesting that AI's positive innovation impact emerges primarily when paired with substantial R&D investment. Additionally, capital expenditures significantly enhance patenting outcomes, reinforcing the need for complementary resources. The findings indicate that AI adoption alone does not guarantee innovation; instead, its effectiveness relies on integrating advanced technologies with robust R&D infrastructure and capital investment. This study contributes to the discourse on AI-driven innovation by demonstrating the contingent nature of AI's impact, emphasizing the importance of aligning AI adoption with strategic R&D and capital investments.

4 - Using Embeddings to Understand in AI Decision Making

Yidan Zhang, Cornell University, Ithaca, NY, United States, X.Y. Han

While recent work in OR/OM increasingly explores /how/ to better train AI algorithms for decision-making, less attention has been given to a complementary question: once AI predictors are successfully trained, what rules has the AI actually learned? This talk demonstrates how the geometry of a network's learned embeddings can shed light on this issue. Using representative OR decision-making problems, such as multi-class scheduling, we show that after an AI learns to solve these problems, its internal vector representations can be mapped to classically understood models. This mapping allows for interpretability in the AI's decision-making procedure.

5 - AI-Assisted Triage of Patient Messages: A Clinically Calibrated Human AI Framework for Healthcare Operations

Mohamed Megahed, The Ohio State University, Columbus, OH, United States, Nathan Craig, Aravind Chandrasekaran

This study addresses clinician burnout from increased asynchronous patient messaging by developing and evaluating an AI-assisted triage system. Leveraging natural language processing and clinician-calibrated severity scoring, the system prioritizes messages for faster, more effective responses. Using data from 3.2 million messages, the research compares four triage workflows, linking severity flags to adverse clinical events. Results show AI-human collaboration improves response time and outcomes while maintaining alignment with clinician judgment. This integrated framework offers a scalable approach to enhancing healthcare operations and advancing human-AI interaction theory.

6 - Resolving AI Value Paradox: Unlocking AI's Value Through Strategic Knowledge Integration

Kaige Gao, Case Western Reserve University, Cleveland, OH, United States

Artificial Intelligence (AI) has transformed industries, yet firms often struggle to extract value from it. To address this AI value paradox, we examine how cross-boundary AI knowledge—AI expertise applied within specific industries—influences a firm value. Using a longitudinal dataset of patents, academic papers, and firm data, we find that how firms integrate cross-boundary AI knowledge with complementary assets significantly impacts their value. Specifically, firms leveraging cross-boundary AI knowledge in exploration strategies are more likely to see higher market returns, while those focusing on exploitation face greater challenges. This study extends the literature on knowledge transfer and exploration-exploitation strategies in the AI era, offering insights into AI's application in the industry and providing practical strategies for firms to maximize AI's value in innovation efforts.

TA64

Building B Level 4 B404

Bridging Theory and Human Behavior in Operations

Invited Session

Behavioral Operations Management

Chair: Arturo Estrada Rodriguez, Católica-Lisbon School of Business and Economics, Lisboa, N/A, Portugal

Co-Chair: Siddharth Singh, University College London, London, E14 5AA, United Kingdom

1 - Non-Truthful Mechanism Design for Revenue Management in Queueing Systems

Arturo Estrada Rodriguez, Católica-Lisbon School of Business and Economics, London, United Kingdom, Siddharth Singh

We investigate how a capacity-constrained firm should define a price/lead-time menu and a scheduling policy in order to maximize revenues subject to the choices of heterogeneous customers with private types (service valuation and delay sensitivity). Traditionally, menu design in such settings has been focused on truthful incentive-compatible (IC) mechanisms, under which rational customers have the incentive to truthfully reveal their private types. In such settings, the need to ensure truthful revelation may incentivize providers to artificially inflate wait times by introducing strategic delay. Motivated by the existence of robust behavioural factors such as lying aversion or strategic naivety, we depart from the canonical model by allowing some proportion of customers to be unconditionally honest - unable/unwilling to misrepresent their private information. We find that the presence of such unconditionally honest customers may give rise to optimal non-truthful (NT) direct mechanisms - where it is optimal to allow some lying to occur in equilibrium - that yield revenue rates higher than any IC mechanism. Moreover, we find that optimal NT mechanisms, in contrast to IC ones, do not resort to delay tactics. We characterize such NT mechanisms, identify when they are optimal, and study their social welfare implications.

2 - Wait-Time Targets and Human Task Scheduling in Queues

Tomas Momesso, UCL School of Management, London, United Kingdom, Rouba Ibrahim, Bilal Gokpinar, Mirko Kremer

How do performance targets influence human scheduling behaviour in service operations? We investigate how wait-time targets (deadlines) can improve task scheduling in high-workload queueing systems. We design a controlled experiment where participants manage a queue with two task types (long and short) and aim to minimise linear delay costs. Tasks arrive dynamically and vary in their delay cost and (exogenously specified) service times. Our results enhance the understanding of how human servers schedule tasks under high workload and how performance targets affect discretionary task ordering.

3 - Transshipment with Overconfidence

Xuan Zhao, Wilfrid Laurier University, Waterloo, ON, Canada, jialu Li

This paper investigates how overconfidence affects the equilibrium of lateral transshipment systems. We find that overconfidence bias can exert a positive force on the decentralized transshipment system, in contrast to its consistent negative effect on a centralized system. Thus, contrary to intuition, overconfident newsvendors can be better off with a decentralized transshipment practice than a centralized one. Additionally, while conventional wisdom holds that coordinating transshipment prices exist to allow the decentralized system to achieve full coordination (first-best outcome), our analysis reveals that such coordination prices fail to exist when overconfidence bias is high. If this is the case, we show that newsvendors with a high (low) net profit margin benefit more from increased (decreased) transshipment prices. Our findings provide actionable insights on how overconfident newsvendors can design transshipment prices more effectively and suggest when to leverage decentralized transshipment practice depending on the newsvendors' overconfidence bias.

4 - Innovation and Information Free-Riding: An Experimental Investigation of Collective Experimentation

Yaroslav Rosokha, Purdue University, West Lafayette, IN, United States, Timothy Cason, Stanton Hudja

We study information free-riding in a dynamic collective experimentation setting. In theory, a social planner can reduce the benefit to free-riding by limiting future access to the innovation. To encourage more experimentation, the optimal amount of access balances the trade-off between too much access that leads to information free-riding and too little access that limits learning. We report a controlled lab experiment that varies future access to the innovation to test several theoretical predictions. Results show that human subjects respond to variation in the future access as predicted; however, the magnitude of their response is insufficient to generate the overall benefit of the intervention. We explore how individual characteristics such as risk aversion, loss aversion, cognitive ability, and ability to update beliefs optimally impact individuals' decisions to information free-ride.

5 - Solver Capacity Utilization and Allocation on Crowdsourcing Platforms: An Experimental Study

Ramazan Kizilyildirim, UCL, LONDON, United Kingdom

We employ a game-theoretic model combined with laboratory experiments to examine how solvers allocate and utilize their capacity across multiple contests as a crowdsourcing platform expands. Our experimental findings reveal that platform growth does not inherently diminish the profit per contest. To further understand these results, we utilize a structural model to analyze the observed data, uncovering the underlying behavioral mechanisms that drive these outcomes.

TA65

Building B Level 4 B405

AI Applications and Ethical Implications

Invited Session

Artificial Intelligence

Chair: Anqi Xu, Bentley University, Waltham, MA, United States

Co-Chair: Ruiyun Xu, Miami University, Oxford, OH, United States

1 - Humanizing the Stay: Impact of Promoting Suppliers in Peer-to-Peer Markets

Yinan Yu, University of Oklahoma, Norman, OK, United States

Peer-to-peer (P2P) markets consist of a wide variety of individual suppliers who often take a back seat to their products. There is an ongoing debate about whether P2P platforms should place suppliers in a more prominent role in their online presence. We study a program promoting hosts in addition to their property offerings on Airbnb, the largest P2P home-sharing platform. In line with the Elaboration Likelihood Model of persuasion, we find that featuring host images on top of their property listings can boost property bookings by 4.2%. Drawing upon a theoretical framework of Stimulus-Organism-Response (S-O-R), we further unveil the mechanism by examining how image attributes (S) and image perception (O) each influence property bookings (R). We leverage a Generative AI approach to extract the objective attributes together with the perception conveyed in the host images. The results show a significant role of image attributes and perception. For example, properties with host images that feature a single person, depict a younger host, have higher resolution and a larger face-to-image ratio, or show a host perceived as friendly are more likely to receive bookings. The impact of host images, however, varies by host and property types. Our findings add to the debate about the online marketing strategy of P2P markets, which are largely driven by individual suppliers and provide novel evidence from Generative AI analytics that informs platform design and supplier management.

2 - Chatgpt vs. Google: a Comparative Study of Search Performance and User Experience

Ruiyun Xu, Miami University, Oxford, OH, United States

The emergence of ChatGPT, a large language model-powered chatbot, has the potential to revolutionize the way users search for information by shifting from keyword-based queries to natural language interactions. This study investigates user behavior differences when using search engines and chatbot tools for various information retrieval tasks and explores the combined use of both tools. We conduct two randomized online experiments to achieve this goal. Our findings reveal that ChatGPT users consistently spend less time on all tasks, with no significant difference in overall task performance between the ChatGPT and Google Search groups. Notably, ChatGPT excels at addressing common information look-ups and providing summarized solutions, but struggles with fact-checking tasks which require users to take extensive exploration to resolve conflicting information. The impact of cognitive fit between task requirements and search tools' design features on performance is task-dependent. Our results also indicate that a strong belief in search result accuracy can diminish the effectiveness of using ChatGPT. When both tools are accessible, the outcomes lie between the results of using each tool individually. Participants strategically employ both tools, often selecting the most suitable one for a specific task. Our study offers valuable insights for future development and integration of chatbot tools with search engines, suggesting potential improvements to enhance user experience and information-seeking efficiency.

3 - Why Do Some Moocs Attract More Subscribers than Others? A Social Presence Theory Perspective

Bo Wen, Northern Arizona University, Flagstaff, AZ, United States, Mandy Dang, Sophie Friedman

As MOOCs (Massive Open Online Courses) continue to transform how people learn new knowledge, an important question arises: why do some courses attract more subscribers than others? Draw on social presence theory, this study investigates two main design features that may drive learners to subscribe to MOOCs: instructor presence in the course cover image and the use of learner-focused course description. Analyzing data from over 3,000 Udemy courses, we find that courses featuring the instructor in the cover image attract more subscribers, even when controlling for other key variables. Additionally, course descriptions that use second-person pronouns (e.g., "you will learn") are more effective in attracting subscribers than those focused on the instructor (e.g., "I will teach"). These effects are robust across different course categories and model estimations. Post-hoc analyses further reveal that the positive effects of both instructor presence and learner-focused descriptions are amplified in more advanced courses. In addition, we find evidence that these two design features could interact positively to increase the number of subscribers. Together, these findings offer new insights into the subscription-based revenue model in MOOCs and suggest AI-driven strategies to enhance learner engagement, such as detecting missing instructor presence and optimizing course descriptions to be more learner-centric.

4 - Modeling Individual Credit Card Consumption Behavior and Delinquency with An LLM-Based Agent

Yu Zhu, University of Delaware, Newark, DE, United States, Bowen Yin, Minjia Mao, Xiao Fang

We propose an agent-based simulation framework that leverages large language models (LLMs) to emulate credit card user behavior, focusing on spending and repayment actions. To realistically model user dynamics, we introduce a novel memory architecture that integrates both user profiles (expressed in natural language) and transaction histories (encoded as graphs), enabling the agent to reason over static traits and behavioral patterns. This architecture features three key innovations. First, we construct two complementary graph representations of transaction history: one that captures individual behavioral patterns, and another that reflects aggregate transaction patterns across the full user population. These graph-based representations better capture interdependencies among transactions compared to traditional sequential formats. Second, we design an adaptor module to align the graph representations with the LLM's embedding space, facilitating seamless integration of structured financial data with natural language understanding. Third, we introduce a reflection mechanism that enables the LLM to dynamically retrieve relevant portions of the transaction graphs during

simulation. This approach allows the foundation LLM's parameters to remain frozen, requiring training only the lightweight reflection module. We demonstrate the utility of our framework in two key applications: delinquency prediction and intervention evaluation. Our experiments show that the LLM-based agent achieves predictive accuracy on par with conventional models in delinquency forecasting. Moreover, the simulated spending and repayment behaviors align closely with expected trends under varying macroeconomic scenarios, underscoring the framework's capacity to generalize under out-of-distribution conditions.

TA66

Building B Level 4 B406

Data-Driven Innovations in OR Education

Contributed Session

Contributed

Chair: Utkarsh Gupta, Birla Institute of Technology

1 - The Pivot Summation Method: A Geometric Reimagination of Simple Linear Regression

Om Agrawal, Columbia University, New York City, NY, United States

In this paper, we will present a geometric re-interpretation of OLS regression for calculating the line of best fit in simple linear regression. This approach is using a novel pivot-based vector summation approach. Instead of creating the line of best fit using algebra by solving normal equations simultaneously, this paper aims to present a idea to consider each data point in the dataset as a pivot point, then construct the directional vector from that pivot point by iterative vector summation which then defines the slope of the line. We then iteratively choose the line that minimizes the root-mean squared error, where the algorithm uncovers the optimal direction. We mathematically prove that under conditions where the data is symmetrical, this approach converges with the classical OLS least squared result. We do this using the first order Taylor expansion around the data centroid. Additionally, we explore how the data asymmetry influences the slope and thus giving insights on a geometric intuition into leverage and influential points. While this paper doesn't propose a computational improvement over the OLS, this framework provides a conceptually rich, mathematically backed and a visually intuitive pathway to understanding regression. This approach can have potential value in statistics education and enhance the interpretability of the fundamental mathematical method to model data. Empirical results on the synthetic dataset supports the theoretical convergence, and visualizations reinforce the method's explanatory power.

2 - Improving Decision Analytics Education Using Team-Based Learning and Group Development

William Millhiser, Zicklin School of Business, Baruch College, CUNY, New York, NY, United States

We redesigned an undergraduate course in decision analytics around two related literatures: team-based learning (TBL) and group development. To our knowledge, this is the first to fully integrate best pedagogical practices in TBL into an entire decision analytics course, with adherence to the science of group development and behavioral anchors in the peer review process. Specifically, classroom time is restructured around team-based problem solving via 25 mini cases written by the author and sequenced to deliver the class learning goals. The students work in teams during class time, with a required team deliverable. Before each mini case, students prepare by watching a 10- to 15-minute "prep" video with accompanying materials; similarly, after submitting each deliverable, students "unlock" a debriefing video and corresponding materials. New theory, tools, and techniques are presented in the prep or in the debriefing, not a textbook. Materials are made available to students for free, part of the university's zero-textbook cost and open educational resources (OER) initiatives. Effective strategies in team-based learning are employed throughout the course: group formation and management, accountability for individual and team work (a team charter, 360-degree peer feedback), team exams, and promoting team learning and team development. We will present the course redesign and its objectives, connect the design to literature and evidence on the effectiveness of team-based learning and group development, and conclude with suggestions for further research.

3 - Rethinking CLV: Organizational Barriers to Adoption and the Future of Customer Value Metrics

Utkarsh Gupta, Birla Institute of Technology, Mesra, India, Satyajit Mahato, Satish Chander

Customer Lifetime Value (CLV) is widely recognized as a key metric in marketing analytics, offering a long-term view of customer profitability. Yet, in practice, its adoption often lags behind to short term KPIs or campaign level metrics. This disconnect between academic consensus and business implementation reveals a niche but critical challenge: why does a theoretically robust metric like CLV struggle to gain traction in business decision making?

This study examines the organizational dynamics influencing CLV adoption, focusing on subscription-based firms where customer value insights are especially relevant. Using semi-structured interviews with analytics and marketing leaders, we examine select enablers and barriers such as model interpretability, incentive alignment, and executive sponsorship, that influence whether CLV becomes embedded in strategic workflows.

Building on these insights, we propose the CLV Adoption Framework (CAF), a diagnostic tool designed to assess business readiness across three dimensions: analytical maturity, decision integration, and strategic alignment. While technical advancements in CLV modeling are well documented, few studies offer a structured, practitioner-focused lens on adoption challenges. The CAF addresses this gap by translating qualitative findings into an actionable roadmap for organizations.

Finally, we raise a forward-looking question: Has CLV, in its current form, reached its practical limit and should firms begin exploring more interpretable, action-oriented customer value metrics? This research contributes both conceptually and practically to marketing analytics and is especially relevant for data scientists, marketing strategists, and business leaders working to bridge the gap between predictive models and real-world impact.

4 - Increased Engagement and Learning Outcomes from the Production Planning Game

Melissa Tilashalski, Johns Hopkins, Baltimore, MD, United States, Kimberly Ellis, Natalie Cherbaka

The Production Planning Game is a web-based, experiential learning, manufacturing game. Teams of students prepare production plans by creating the aggregate plan, master production schedule, and material requirements plan for a set of products and then face off with other teams through simulated demand across nine time periods to earn the most profit in order to win the game. Effective experiential learning activities increase student engagement, which is considered a strong predictor of student learning. With the recent increase in distance learning, instructors are interested in experiential learning activities that are effective for multiple modalities. This research had teams of students play a web-based educational game, The Production Planning Game, either in-person or remotely, as a culminating learning experience in a required junior level course in industrial and systems engineering. Student engagement and learning outcomes were evaluated across the two modes of delivery, in-person and remote. The online experiential learning activity is found to be effective at improving student engagement in three ways - cognitively, emotionally, and physically - and at improving student learning outcomes, regardless of modality. This talk will present the results of the study as well as explain how to implement the experiential learning activity - The Production Planning Game - in your course.

5 - Sequential University Admission Optimization with ML-Based Enrollment Forecasts

Meiling Wu, University of California, Irvine, Irvine, CA, United States

University admission decisions are made across multiple rounds under uncertainty about applicants' future enrollment behavior. In this work, we develop a sequential admission optimization framework that integrates machine learning-based enrollment probability predictions with a dynamic optimization model inspired by the newsvendor problem. Our objective is to help universities meet enrollment targets while balancing trade-offs among class size, academic quality, and scholarship budget usage.

We first apply supervised learning models to estimate applicant-level enrollment probabilities using historical admission and enrollment data. These individualized forecasts are then used as inputs to a sequential decision model that determines whom to admit or waitlist in each round of the admission cycle. The optimization

framework extends the classical newsvendor logic by incorporating asymmetric costs of over- and under-enrollment, and it is solved via dynamic programming to account for future rounds and applicant pools.

Unlike traditional models that optimize admissions in a static, single-shot setting, our approach explicitly handles the dynamic nature of admissions and the delayed realization of enrollment outcomes. The final enrollment outcome is only observed at the end of the admission cycle, making the problem resemble a delayed-reward optimization. We show that our model improves decision quality over myopic or deterministic heuristics by anticipating future applicant inflows and reacting to stochastic applicant behavior.

6 - Adaptive Convoy Logistics: A DDDA Framework with Dynamic Bayesian Belief Networks

Elnaz Alinezhad, Wayne State University, Detroit, MI, United States

This work presents a Dynamic Data-Driven Applications System (DDDAS) for adaptive fidelity management in digital twins, with a focus on convoy logistics in contested environments. A Dynamic Bayesian Belief Network (DBBN) is developed to estimate latent system states using multimodal sensor observations, enabling real-time probabilistic reasoning under uncertainty. The latent variables modeled include Environment, Vehicle, Convoy, Operation, Human, and Enemy states, each inferred from heterogeneous inputs such as terrain, visibility, fuel, maintenance, health, and risk indicators. The system adaptively adjusts sensor fidelity levels and model update rates based on Bayesian posterior probabilities, minimizing computational load while preserving decision-relevant accuracy. A momentum-based prior update mechanism, controlled by a smoothing parameter α , is incorporated to balance responsiveness and belief stability. Through sensitivity analysis across varying α -values, we demonstrate how fidelity dynamics respond to memory-depth trade-offs. Simulation results show that environmental and vehicle subsystems exhibit sustained high-fidelity periods, while human and adversarial nodes trigger more transient responses. This framework enables scalable, resource-efficient digital twins that remain operationally responsive, particularly when integrating computationally expensive human-system components.

7 - Modeling the Evolution of Driver Trust in Conditionally Automated Vehicles Using a Bdi-Based Dynamic Bayesian Framework

Roohollah Jahanmahin, Wayne State University, Detroit, MI, United States, Sara Masoud

This research presents a cognitively inspired framework for modeling how driver trust evolves in conditionally automated vehicles. Using the Belief–Desire–Intention (BDI) model embedded in a Dynamic Bayesian Network (DBN), trust is represented as a latent state influenced by system performance, driver goals, and emotional context. Belief reflects the driver's perceived AV reliability, desire captures motivation to delegate control, and intention models the resulting behavioral decision. The model combines subjective trust ratings with observable behavior (e.g., gaze, AV outcomes) to infer and predict trust over time. A utility-based probabilistic function is trained to estimate intention decisions, enabling real-time trust monitoring. This framework supports the development of adaptive, trust-aware AV systems that can respond to individual cognitive states and behavioral changes.

8 - How Digitization Shapes Scientific Collaboration? Evidence from Electronic Health Records

Max Wang, Columbia Business School, New York, NY, United States

Digitization transforms knowledge production in teams by improving data interpretability and shareability. This paper validates this claim by estimating the effect of adopting Electronic Health Records (EHR) on scientists working at US hospitals. EHR is a digitization technology that standardizes and centralizes patient record data. It is designed to improve care delivery, and its adoption has been regulated since 2009. Using a panel of scientists from 148 teaching hospitals, I show that EHR adoption increases collaboration by 6.1% and broadens intellectual diversity by 1.1%. These effects surpass those of adopting conventional health information technologies without data standards and intensify when collaborators share a common EHR system. I then show that EHR adoption compromises research outcomes in citations and novelty, likely due to adjustment costs. Senior scientists, whose experience attenuates adjustment costs, benefit more from working with EHR. I corroborate my analyses using interviews with real-life hospital scientists.

9 - Optimizing Speed and Carbon in Omnichannel Fulfillment: AI-Powered Next-Day Delivery & Same-Day Pickup

Nixalkumar Patel, LG Electronics, Lake Hiawatha, NJ, United States

In the face of rising customer expectations and sustainability mandates, a large omnichannel enterprise reimagined fulfillment with advanced AI and analytics. In this session, we will demonstrate how to build ZIP-level demand-forecasting models and deployed intelligent allocation algorithms to optimize both next-day home delivery and same-day in-store pickup. Attendees will see how these techniques increased conversion and fulfillment reliability, while delivering measurable reductions in last-mile carbon emissions. We will examine key AI and machine-learning methods, explore the environmental trade-offs of AI-driven routing and delivery-window management, and present a scalable framework applicable across industries. You will leave with a practical blueprint for designing and implementing smarter, greener fulfillment strategies at scale.

TA67

Building B Level 4 B407

Data Mining for Inference and Mathematical Conjecturing

Invited Session

Data Mining

Chair: Yanjun Qian, Virginia Commonwealth University, Richmond, VA, United States

1 - Predictive Inference for Ion Concentration Using Ensemble Models

Yanjun Qian, Virginia Commonwealth University, Richmond, VA, United States

Accurately measuring trace amounts of Nickel (Ni) and Cobalt (Co) is crucial for environmental monitoring, assessing toxicological risks, and ensuring quality control in industrial processes. While UV-Vis spectroscopy offers a rapid and non-destructive alternative to traditional chemical assays, it presents its own statistical challenges, including high dimensionality, significant multicollinearity, nonlinear signal-response relationships, and often limited sample sizes. We propose utilizing a stacked ensemble approach for spectral regression tasks, which consistently delivers stronger predictive accuracy compared to its individual components. Additionally, we incorporate uncertainty quantification methods to provide well-calibrated and interpretable confidence intervals, making this framework suitable for regulatory and decision-making contexts. Overall, this work presents a robust and principled pipeline for predicting ion concentrations, leveraging insights from uncertainty quantification.

2 - Spatiotemporal Machine Learning Framework for Predicting Dengue Incidence in the Dominican Republic

Sahil Chindal, Virginia Commonwealth University, Richmond, VA, United States, Demian Herrera, Manuel Colome-Hidalgo, Helena Sofia Rodrigues, Yanjun Qian, Michael Robert

Dengue is a mosquito-borne disease endemic to many tropical and subtropical regions around the world, including the Dominican Republic. Temperature, precipitation, humidity, and other climate variables play a crucial role in dengue spread due to their association with mosquito growth and proliferation. Using climate and dengue case data, we develop models for spatiotemporal analysis of dengue transmission between various provinces of the Dominican Republic. We estimate the reproduction number (R_0) and effective reproduction number (R_t) for each province using traditional statistical methods as well as various ensemble learning and deep learning models. We train our model for multiple years of data, and compare the accuracy of our methods to the expected values of R_0 and R_t for dengue in the region. We investigate graph neural networks (GNNs) to explore the relationship of dengue case transmission between provinces. We aim to construct a model that can help determine the flow of dengue transmission between provinces. Our model can be integrated with geographic, sociodemographic, and other types of information as part of a comprehensive early warning system that predicts outbreaks and informs public health and mosquito control policies.

3 - Reinforcement Learning for Mathematical Conjecturing

Samuel Paul Gyamfi, Virginia Commonwealth University, Henrico, VA, United States

Mathematical conjecturing plays a foundational role in scientific discovery, offer-

ing interpretable expressions that describe or bound relationships between variables.

Existing tools like the CONJECTURING framework, developed by Brooks et al., automate this process using the Dalmatian heuristic—a deterministic approach that generates symbolic expressions as trees and selects those that are valid across all data and not outperformed by others. While powerful, this heuristic is static: it does not learn from past experience, struggles with noisy or incomplete data, and lacks flexibility in balancing complexity and accuracy.

This project proposes a new direction: replacing the Dalmatian heuristic with a Reinforcement Learning (RL) framework. Rather than relying on fixed rules, an RL agent would learn to evaluate and filter expressions based on reward signals that account for both approximation error and structural simplicity. Each candidate expression becomes a decision point—should the agent keep or discard it? Over time, the agent would learn a generalizable policy for selecting high-quality, interpretable expressions tailored to the data.

4 - Data-Driven Analysis of Arctic Vessel Navigation: Mining Historical Shipping Patterns for Navigational Behavior Inference and Risk Assessment **Mauli Pant, Virginia Commonwealth University, Richmond, VA, United States, Linda Fernandez**

The Arctic Northwest Passage (NWP) offers emerging maritime opportunities, but also considerable risks due to shifting environmental and ecological conditions. This study analyzes Automatic Identification System (AIS) data from 2010 to 2019 to evaluate vessel speed as a proxy for backcasting navigational decisions under varying Arctic conditions. Key environmental drivers—sea ice concentration, wind speed, bathymetry, and whale distribution—are examined for their impact on vessel behavior and navigational risk.

Missing data on sea ice, wind, and bathymetry are interpolated using spatio-temporal kriging. Generalized Linear Mixed Models (GLMMs) with a Gamma distribution and log link are used to model Speed Over Ground (SOG), addressing zero-inflation and heteroscedasticity. The models incorporate spatial and temporal random effects to account for geographic and seasonal variability in vessel behavior.

The overlap between vessel tracks and environmental data enabled assessment of ice-related risks using Arctic Climate Change Economy Society (ACCESS 2015) safe speed thresholds. Findings show that cargo and tanker vessels frequently exceed these thresholds, with approximately 40% of vessels from 2010 to 2019 exposed to ice-related risks. GLMMs explain 67% of SOG variability compared to the null model, with spatial effects alone accounting for 31% of the variance. Interactions between vessel group and ice concentration reveal how vessels adjust speed in response to ice. Similarly, interactions between vessel status and bathymetry highlight how navigation decisions are influenced by ocean depth.

Future work will incorporate whale distribution data to evaluate ship strike risks, contributing to a risk optimization framework for sustainable Arctic maritime operations.

5 - Genotype Classification Using High-Resolution Melt Curves and Machine Learning Approaches **Parastoo Bybordi, Virginia Commonwealth University, Richmond, VA, United States, Indranil Sahoo**

In forensic science, high-resolution melt (HRM) curve analysis, a method for DNA sample analysis, plays an important role in criminal cases. High-resolution melt (HRM) curves provide valuable insights that help investigators identify people based on their DNA profiles at crime scenes. This is critical in situations where the DNA samples are degraded or exist in low amounts, which makes the analysis more challenging. Moreover, High-resolution melt (HRM) analysis is a cost-effective and rapid technique to confirm or rule out individuals from investigations by comparing the genotypes of DNA samples found at the scenes. HRM curves provide predictions about genotypes with accurate results even with degraded DNA samples. Machine learning methods are valuable tools for accurately analyzing HRM curves and improving the prediction and classification of genotypes from DNA samples. Machine learning algorithms precisely extract features from high-resolution melt (HRM) curves to identify and categorize the genotypes. Using these advanced methods in the initial stage of DNA analysis before the sample becomes degraded can increase the accuracy of identifying contributors' genotypes. This investigation aims to develop machine learning techniques that enhance the accuracy and reliability of genotype classification in scenarios where other DNA analyses, like STR profiling, may be less effective due to a limited amount of DNA sample data.

6 - Spatiotemporal ML Prediction of Egg Counts of the Dengue Vector *Aedes Aegypti* in Cordoba, Argentina **Sushyanth Sridhar, Virginia Tech, Blacksburg, VA, United States, Robert Michael, Esra Toy**

Abstract:

Dengue fever has escalated to critical levels across Latin America, with over 13 million cases and 8,000 deaths reported in 2024 alone. Accurate forecasting of *Aedes aegypti* populations is essential for proactive vector control. This study proposes a spatio-temporal modeling framework to predict weekly egg counts using ovitrap surveillance data in Córdoba, Argentina. We evaluate machine learning (ML) models—Random Forest, Gradient Boosting, XGBoost—and LSTM over a 28-week breeding season (November–May). Historical data from 178 traps (2009–2013) and recent records from 33 traps (2023–2024) form the observational basis. To address limited recent surveillance, we introduce a novel data augmentation method using K-means clustering and random resampling, validated via Fourier analysis to preserve temporal frequency characteristics. ML models trained on real ovitrap data achieved R^2 scores above 0.95, with Random Forest consistently outperforming other ML methods. An LSTM model trained solely on synthetic data also demonstrated high predictive fidelity. Environmental features, including LANDSAT-7 land cover indices and NASA POWER weather data, served as critical inputs. Our results reveal strong associations between mosquito incidence and urbanization, surface moisture, and small water bodies, with reduced presence in arboreal zones. Peak egg counts consistently occurred after week 16. This framework offers scalable, data-efficient strategies for guiding vector control in resource-constrained environments.

TA68

Building B Level 4 B408

Machine Learning and Optimization for Decision-Making Under Uncertainty

Invited Session

Data Mining

Chair: Dervis Ozay, The University of Texas at Arlington, Arlington, TX, United States

Co-Chair: Shouyi Wang, The University of Texas at Arlington, Arlington, TX, United States

1 - Design Optimization Techniques Under Uncertainty, Incorporating Online Learning and Calibration to Account for Evolving Market Conditions

Shouyi Wang, The University of Texas at Arlington, Arlington, TX, United States

Dynamic pricing under uncertainty is critical for small and medium-sized sellers operating in competitive e-commerce markets. Existing platform tools, such as those offered by Amazon, often provide only basic rule-based adjustments—setting minimum and maximum thresholds or staying below competitors' prices—without accounting for seasonality, customer behavior, or broader competitive dynamics. This study develops an adaptive optimization framework that combines probabilistic demand forecasts with techniques such as Bayesian optimization, grid search, and random search to determine optimal pricing and advertising expenditure. By integrating online learning and calibration, the framework adapts to evolving market conditions while remaining computationally efficient and accessible for smaller sellers. Validation in simulated market environments shows that the approach improves pricing precision, captures demand fluctuations, and supports practical, affordable dynamic pricing strategies beyond the limitations of current platform tools.

2 - Enhancing Healthcare Access in Rural Appalachia via a Hybrid Truck&Ndash;Drone Delivery System Powered by Deep Reinforcement Learning

Sabrullah Deniz, The University of Tennessee Knoxville, Knoxville, TN, United States

Timely delivery of critical medical supplies in rural and mountainous regions is hindered by poor infrastructure, long travel distances, and unpredictable conditions. This work proposes and evaluates a hybrid truck–drone delivery framework optimized using multi-agent deep reinforcement learning (MADRL) to improve healthcare logistics in such challenging environments. Trucks act as mobile depots transporting bulk supplies to strategically chosen stops, from which drones execute agile last-mile deliveries. The operational decision-making problem is modeled as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP) and solved using a centralized-training, decentralized-execution (CTDE) MADDPG architecture with an aggregator-based critic, enabling scalable coordination under partial observability. We develop a custom simulation environment for East Tennessee's rural topology, incorporating realistic fuel, battery, traffic, and weather constraints. Experiments compare single-truck and multi-truck fleet configurations, each paired with varying drone fleet sizes. Results demonstrate that (i) integrating drones significantly reduces last-mile delivery times and truck mileage, (ii) increasing the number of trucks enables workload partitioning that lowers both delivery time and operational cost, and (iii) an optimal fleet balance exists where the marginal benefit of additional drones diminishes. The proposed DRL-driven hybrid system achieves up to 30–40% reduction in total delivery time and 20% cost savings compared to traditional truck-only logistics. This framework is adaptable to other rural and underserved regions, providing a scalable, data-driven approach to improving healthcare accessibility.

Keywords: rural healthcare logistics, hybrid truck–drone delivery, deep reinforcement learning, multi-agent systems, last-mile delivery optimization

3 - Energy Infrastructure Planning Under Extreme Weather Uncertainties

Thushara Gunda, Sandia National Laboratories, Livermore, CA, United States

Extreme weather poses significant risks to critical infrastructure operations. However, the uncertainties associated with these events pose even more challenges to energy planning activities that are looking out 5 to 50 years into the future. This talk will summarize multiple efforts that are aiming to leverage machine learning and advance data analytics to support characterization of current impacts at both the generator and at grid scales. We will also describe how these impacts are being projected into the future to support identification of optimal investment decisions.

4 - How Uncertain are Your Confidence Intervals: An Investigation on Molecular Properties Prediction

Priscilla Avegliano, IBM Research, São Paulo, Brazil, Emilio Vital Brazil, Eduardo Soares, Victor Shirasuna

Uncertainty is an inescapable part of any decision-making process that takes place in the real world. Even state-of-the-art Machine Learning Models are susceptible to errors that can stem from a myriad of uncertainty sources, ranging from measurement errors to model bias. To tackle this issue, confidence intervals are often used to express how certain a model is about its predictions, but these intervals are themselves error-prone, and if they are over or underestimated they can lead to poor decisions. Hence, the calculation of confidence intervals must be subject to a careful analysis to maximize its contributions. Conformal Prediction is a promising method for generating confidence intervals. It works well with small datasets and makes minimal assumptions about the data or model. A key idea in this method is to estimate how different a new input is from previously seen data. This difference, expressed by a “distance” measure, helps determine how wide the confidence interval should be—the more unfamiliar the input, the wider the interval. In this work, we investigate whether measuring distance in an embedding space, where data is represented in a compressed and more structured form, can improve the confidence interval calculation. We conduct experiments using an open-source deep-learning model for molecular properties prediction. Our results compare embedding-based distances to the traditional Tanimoto distance, a common similarity measure in chemistry. We evaluate which approach produces more reliable confidence intervals for decision-making associated with molecular property prediction tasks.

TA69

Building B Level 4 B409

Product and Process Innovation in Healthcare

Invited Session

Technology, Innovation Management and Entrepreneurship

Chair: Zhili Tian, Univ. of Houston, Houston, TX, United States

1 - Toward Fair and Informed Vaccine Allocation: Data-Driven Prioritization of Vaccine Proposals Using Multi-Modal Fusion, Temporal Processing, and Multi-Criteria Decision Optimization

Ashutosh Kumar, Rochester Institute of Technology, Rochester, NY, United States, Ruben Proano

Equitable vaccine distribution in low- and middle-income countries (LMICs) remains a persistent challenge amid financial, logistical, and data limitations. Each year, the Global Alliance for Vaccines and Immunization (GAVI) must allocate a limited budget across diverse country proposals to maintain routine immunization, introduce new vaccines, or respond to outbreaks. Existing decision frameworks often struggle to account for heterogeneous, incomplete data spanning epidemiological, socio-economic, and political domains, resulting in suboptimal and inequitable outcomes.

We propose an AI-driven decision-support system to enhance transparency and contextual relevance in GAVI's annual vaccine funding decisions. The system integrates: (1) a **data imputation module** for reconstructing missing epidemiological data using static, temporal, and latent variables; (2) an **interpretable multimodal fusion pipeline** that learns joint representations across health, infrastructure, and economic indicators; and (3) a **multi-objective optimization model** that prioritizes proposals by balancing health impact, equity, and feasibility under real-world constraints (e.g., cold-chain capacity, limited supply).

We hypothesize that integrating robust imputation and multimodal learning will yield more accurate, context-aware funding recommendations, and that a multi-objective optimization framework will outperform existing rule-based or single-objective methods. Validation will use public (WHO, UNICEF, GAVI) and simulated datasets to assess imputation accuracy, representation quality, and allocation performance, measured via Pareto efficiency, DALYs averted, and equity indices. This work contributes the first unified AI pipeline for vaccine budget allocation, potentially supporting more equitable and effective global immunization strategies.

2 - Using Children's Games to Illustrate Lean Principles in Healthcare Process Improvement

Murray Côté, Texas A&M University, College Station, TX, United States

Lean management has become a widely adopted framework for healthcare quality improvement due to its emphasis on waste reduction, standardization, and continuous process improvement. Despite its benefits, widespread and sustained implementation in healthcare settings can be perennially difficult, often due to limited stakeholder engagement and the complexity of fostering a lean mindset. This presentation uses children's games as an educational and validation tool to enhance stakeholder

understanding and application of lean principles in healthcare process improvement. The games *Cootie* and *Hi Ho! Cherry-O* serve as an analog for healthcare processes as the simplicity of the games, combined with the intuitive underlying Markov structure, allows stakeholders to visualize process variability, understand performance metrics such as cycle time, and identify areas for process improvement. Four scenarios are compared to demonstrate how baseline conditions and probability distributions affect overall performance. This approach provides a user-friendly approach that encourages stakeholder engagement and demystify lean methodologies. In turn, our approach contributes a novel, accessible tool for lean education in healthcare, reinforcing the need for stakeholder-driven, experiential learning to ultimately achieve sustainable quality improvement.

3 - Port Congestion and Resilience Effects on Supply Chains

Elham Zafarani, Arizona State University, Tempe, AZ, United States

We examine US port congestion and resilience, focusing on the impact of external shocks, such as weather events, on supply chain flows. We examine these flows over time and across industries, aiming to inform policy on infrastructure investment to alleviate port congestion and improve resilience.

Tuesday, October 28, 9:45 AM - 10:45 AM

Building A Level 4 Auditorium

Plenary - Delta at 100: Harnessing a Century of Insight to Power the Future of Travel with AI

Plenary Keynote

Plenary

1 - Delta at 100: Harnessing a Century of Insight to Power the Future of Travel with AI

Eric Phillips, Delta Airlines, Atlanta, GA, United States

As Delta Air Lines celebrates its centennial year, we reflect on a century of transformation fueled by data, innovation, and a deep commitment to customer experience. From early operational metrics to today's real-time, AI-powered decision-making, Delta's journey highlights how data and insights have continually shaped our strategy and elevated our brand.

This session explores how Delta's digital and AI evolution is not a recent shift, but a natural extension of our longstanding focus on operational excellence and customer-centricity. AI is now embedded across our enterprise—from optimizing crew and aircraft scheduling to delivering personalized, real-time support through the Fly Delta app and emerging tools like Delta Concierge.

These capabilities are redefining how employees embody Delta's values, how we manage global operations with precision, and how we create seamless, tailored experiences for over 200 million customers annually.

Attendees will gain insight into how Delta is leveraging AI to:

Enhance operational reliability through predictive analytics.

Empower employees with intelligent tools that streamline service.

Drive loyalty through hyper-personalized engagement strategies.

Build a scalable digital platform that connects data, people, and purpose.

Join us to discover how Delta's century of experience is powering the next era of intelligent, human-centered air travel.

Tuesday, October 28, 11:00 AM - 12:15 PM

TB01

Building A Level 3 A301

The Ferryfield Group/AMPL

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - AI-Augmented Monte Carlo Simulation for Capital Investment Decisions in Innovation Project Portfolios

Gustavo Vinueza, Ferryfield, Ithaca, NY, United States

Capital investment decisions in innovation project portfolios are characterized by high uncertainty—ranging from technical feasibility and development timelines to market adoption and regulatory approval. Traditional deterministic models often fail to capture these uncertainties, leading to suboptimal capital allocation and increased risk exposure. This paper introduces a decision-support framework that combines Monte Carlo Simulation (MCS) with artificial intelligence (AI) to support more robust, data-informed investment strategies. Monte Carlo Simulation is used to model uncertainty across financial and scheduling dimensions, producing probabilistic estimates for key metrics such as Net Present Value (NPV), Internal Rate of Return (IRR), and the likelihood of cost and time overruns. To improve simulation fidelity, AI techniques—particularly machine learning—are applied to historical and contextual project data. These models help refine input distributions, detect interdependencies between projects, and dynamically adjust assumptions in response to new evidence. Beyond input calibration, AI is used to support optimization efforts by analyzing simulated outcomes and identifying portfolio strategies that maximize expected value under defined risk constraints. This enables decision-makers to explore trade-offs between risk and return, better allocate limited capital, and gain visibility into the impact of uncertainty drivers. The proposed framework is demonstrated through a fictitious innovation portfolio composed of diverse early-stage projects. Results show improved transparency, better risk-adjusted returns, and actionable insights for capital allocation under uncertainty. The methodology is applicable to industries where innovation drives value, such as pharmaceuticals, energy, and technology.

2 - From Classroom to Industry: Modern Optimization with AMPL for Energy, Finance, Supply Chain, and Beyond

Gleb Belov, AMPL Optimization Inc., Melbourne, Australia, Marcos Dominguez Velad

Optimization has never been more critical – or more accessible. This talk will show how students and applied researchers can bridge the gap between academic learning and real-world impact using AMPL in modern workflows. With free academic licenses now including commercial-grade access to leading linear solvers (Gurobi, CPLEX, Xpress, Mosek, COPT, HiGHS and more), AMPL offers an unmatched opportunity to develop and apply optimization skills directly in industry contexts. We will explore how AMPL integrates seamlessly with the tools you already know – Python, R, Jupyter, Colab, VS Code, cloud platforms, and modern data pipelines – making it a “set it and forget it” system for building models that stay faithful to the real problems you're solving. Attendees will see how AMPL powers large-scale, mission-critical applications in energy, finance, transportation, and supply chain, enabling companies to save costs, increase efficiency, and improve decision-making. By combining intuitive modeling with cutting-edge solver technology, AMPL empowers students and researchers not just to learn optimization, but to contribute immediately to high-impact, industry-grade projects. Join us to discover how AMPL can position you, or your students, at the forefront of applied optimization.

TB02

Building A Level 3 A302

Solverminds

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Solverminds' Scalable Optimization-as-a-Service**Mukund Sharma, SOLVERMINDS, Alpharetta, GA, United States, Senthil Kumar**

Explore how Solverminds is redefining optimization through client-centric, co-developed solutions. This session highlights use cases across various scheduling challenges, demonstrating real-time decision support, fuel optimization, and scheduling efficiency across the maritime and transport industries.

2 - Solverminds' Innovation in Progress: Optimization for the Future**Mohit Oberoi, SOLVERMINDS, Alpharetta, GA, United States, Senthil Kumar**

This session offers a forward-looking perspective on how Solverminds is developing next-generation optimization solutions. We'll share key concepts, and early-stage innovations from ongoing projects, showcasing how our evolving approach can be applied across shipping, logistics, and broader transport ecosystems.

TB03

Building A Level 3 A303

Navigating Washington: Science, Strategy, and the Shifting Landscape of Federal Policy

Invited Session

Advocacy

Chair: Jeffrey M. Cohen, INFORMS, 5521 Research Park Drive, Baltimore, MD, 21228, United States

1 - Navigating Washington: Science, Strategy, and the Shifting Landscape of Federal Policy**Charles Cooper, Brumidi Group, Washington, DC, United States**

One year after a presidential election—and one year before pivotal midterms—Washington is immersed in one of the most volatile and consequential eras in recent memory. Federal science budgets face mounting pressure, regulatory priorities are evolving, and national debates on AI, supply chains, public health, and national security are intensifying. Join INFORMS Chief Strategy Officer Jeff Cohen and Charles Cooper, founding partner of the Brumidi Group, for a fast-paced, insider's conversation on what's happening in Washington, what's at stake for the operations research and analytics community, and how INFORMS is shaping the policy landscape. This session will unpack the latest developments in federal funding, S&T policy, and workforce legislation—connecting them to your research, your profession, and your ability to drive real-world impact. Whether you're new to policy or already engaged with INFORMS' advocacy efforts, this session will help you cut through the noise—and see where your voice and expertise are needed most.

2 - Navigating Washington: Science, Strategy, and the Shifting Landscape of Federal Policy**Jeffrey M. Cohen, INFORMS, Baltimore, MD, United States**

One year after a presidential election—and one year before pivotal midterms—Washington is immersed in one of the most volatile and consequential eras in recent memory. Federal science budgets face mounting pressure, regulatory priorities are evolving, and national debates on AI, supply chains, public health, and national security are intensifying. Join INFORMS Chief Strategy Officer Jeff Cohen and Charles Cooper, founding partner of the Brumidi Group, for a fast-paced, insider's conversation on what's happening in Washington, what's at stake for the operations research and analytics community, and how INFORMS is shaping the policy landscape. This session will unpack the latest developments in federal funding, S&T policy, and workforce legislation—connecting them to your research, your profession, and your ability to drive real-world impact. Whether you're new to policy or already engaged with INFORMS' advocacy efforts, this session will help you cut through the noise—and see where your voice and expertise are needed most.

TB04

Building A Level 3 A304

ESG, Climate Finance & Sustainability

Contributed Session

Contributed

Chair: Aparna Gupta, Rensselaer Polytechnic Institute, Troy, NY, United States

1 - Estimating GHG Emissions Using Advanced Industrial Classification Systems**Rei Yamamoto, Keio University, Kohoku-ku, Japan, Akifumi Isogai, Masatoshi Nozaki, Kosuke Ishikawa**

This study proposes a novel methodology for estimating corporate greenhouse gas (GHG) emissions using a highly detailed industry classification system. Conventional estimation models often rely on coarse, single-segment industry classifications, which fail to capture the complexity of modern business structures—particularly when estimating Scope 3 emissions. In contrast, we utilize the FactSet Revere Business Industry Classification System (RBICS), which categorizes companies into over 1,800 detailed business segments based on revenue composition. By combining RBICS data with company financials and disclosed emissions data, we construct an Ordinary Least Squares (OLS) regression model that incorporates multi-segment business structures. Our empirical analysis of 4,000 global companies demonstrates that more granular classifications (L6 level) significantly improve the explanatory power of the model compared to traditional methods. Additionally, companies with diversified business portfolios show lower estimation errors, indicating the model's robustness. Unlike machine learning approaches, our method maintains high interpretability, allowing stakeholders to identify which business activities drive emissions. This transparency is particularly valuable for financial institutions aiming to manage climate-related risks and engage with investee companies. Our findings suggest that advanced product-based classification systems, combined with transparent modeling, can meaningfully enhance the accuracy and usability of GHG emissions estimates in sustainable finance.

2 - When the Weather Hits the Market: the Financial Impact of Climate Extremes on Firms in the Energy Sector**Thomas Purchas, University of Washington, Bothell School of Business, Bothell, WA, United States, Xiahua Wei**

This study investigates the potential causal relationship between extreme weather events and the financial performance of U.S. energy sector firms. We focus on the growing occurrence of climate-related disruptions, including hotter summers, colder winters, and a rising frequency and intensity of hurricanes, and analyze their impact on firm's financial outcomes, such as revenues and stock prices. Analyzing historical weather and financial data, we identify patterns of abnormal returns associated with these events. Our findings provide insights to inform investor strategies and enhance understanding of climate-related financial risk in the energy sector.

3 - Disaster Risks and Intergenerational Equity**Heqing Huang, National University of Singapore, Singapore, Singapore, Harrison Hong, Neng Wang**

Disasters, especially ones worsened by climate change, pose a threat to intergenerational equity. As society learns about these disaster risks with unfolding extreme events, we model how to optimally ensure intergenerational equity, which is captured by a constraint that the welfare of future generations be in expectation non-decreasing over time. The planner's solution under this sustainability criterion can be decentralized by disaster-risk state-contingent fiscal policies. When disaster risk is high, the sustainability constraint binds and consumption taxes are then needed to subsidize investments for future generations. Calibrating to the US data, we solve for the optimal fiscal policies and show how intergenerational equity differs with and without the sustainability considerations.

4 - Investment Decision Support for Circular Economy Portfolio Allocation by Firm and Industry**Aparna Gupta, Rensselaer Polytechnic Institute, Troy, NY, United States**

There is an increasing emphasis the world over for reducing waste and making the industrial economy more circular and geared towards sustainability. However, the appropriate advisory for making investment allocation decision to support this development and advancement is significantly missing. In this paper, we develop an

investment analysis framework to support investors' assessment of industries and firms to make their investment allocation decision.

TB05

Building A Level 3 A305

Service Science Award Session IV

Award Session

Service Science

Co-Chair: Renyu Zhang, The Chinese University of Hong Kong, 12 Chak Cheung Street, Hong Kong, N/A

Co-Chair: Muge Yayla Kullu, University of Texas at San Antonio, San Antonio, TX, United States

1 - How Not to Overpackage: AI for Sustainability in Hellofresh's Service Supply Chains

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

Packaging waste is a common challenge faced by meal kit subscription services. Before shipping each meal kit box, providers must decide the grade of liners and the number of freezer packs to include. While these materials are essential for preserving the quality and safety of meal kits in transit to customers, excessive use leaves a larger carbon footprint. As the leader of the meal kit industry, HelloFresh is committed to furthering the sustainability of its supply chain and decreasing its environmental footprint. This paper formulates the contextual packaging problem as an online learning problem, and develops a contextual bandit-based AI method to mitigate overpackaging efficiently. Our algorithm, Contextual One-Sided Arm Elimination, takes into account weather, conditions of transit, box contents and other information to adaptively make the packaging decision for each box. We extract structural information from offline delivery data to accelerate online learning. Our algorithm achieves theoretically optimal performance with an $\sim O(T^{1/2})$ regret bound where T is the horizon length. We test our algorithm on real delivery data from HelloFresh and identify points of overpackaging in the current guidelines. In comparison with the current practice of conducting physical chamber experiments to build packaging guidelines, our algorithm refines the packaging guidelines more swiftly, with higher granularity and lower cost. Our method is effective in identifying overpackaging in existing guidelines, and could smoothen the transition process to new packaging options that are more environmentally friendly in the future.

2 - Achieving Rawlsian Justice in Food Rescue

Aydin Alptekinoglu, Penn State, University Park, PA, United States, Gerdus Benade

We study a problem faced by a national food rescue platform that matches each donation to the first recipient who claims it. Recipients have very different response rates, leading to a few highly responsive recipients claiming the bulk of the donations. We ask whether priority lists, which control when the donation is announced to each recipient, are a remedy for such inequitable outcomes. We show that an n -stage priority list, with an individual notification time specific to each eligible recipient, can achieve any desired expected allocation, hence any fairness target, if the donation is non-perishable. Two-stage or binary priority lists, which notify eligible recipients in only two waves, are simpler to implement and administer but offer less fine-grained control over fairness outcomes. For both perishable and non-perishable donations, we give polynomial-time algorithms to find the n -stage and binary priority lists that optimize a class of Rawlsian objective functions-maximize the minimum value-representing a notion of fairness that focuses on the worst-off. The key insight that animates optimal priority list design throughout is to give higher priority (and thus more time to claim a donation) to recipients who have received less in the past and to those who were slower in responding to notifications. We prove that this idea can be codified into a simple index by which to rank order eligible recipients. Computational experiments calibrated by real data confirm that even simple, binary priority lists lead to significantly more fair allocations than the existing first-come-first-serve allocation system.

3 - Improving Access to Essential Medicines in Sierra Leone via Machine Learning

Angel Tsai-Hsuan Chung, The Wharton School, UPenn, Philadelphia, PA, United States, Jatu Abdulai, Patrick Bayoh, Lawrence Sandi, Francis Smart, Hamsa Bastani, Osbert Bastani

A critical challenge in healthcare systems in Low- and Middle-Income Countries (LMICs) is the efficient and equitable allocation of scarce resources, particularly essential medicines. This problem is complicated by limited high-quality data, which restricts the applicability of traditional data-driven techniques. We propose a novel machine learning framework for essential medicines allocation, which leverages a combination of multi-task learning and decision-aware learning to improve sample efficiency and ensure equitable allocation. In collaboration with the Sierra Leone national government, our framework has been deployed nationwide as a decision support tool to help reduce waste and improve essential medicines allocation. Our evaluation using synthetic difference-in-differences analysis demonstrates a 19% increase in medicine consumption, with no changes to the supply, improving access for approximately 3.7 million women and children under five. Through experimental validation, we demonstrate that our approach also significantly outperforms baseline approaches. Our work demonstrates the tangible impact of machine learning in optimizing high-stakes decisions in resource-constrained settings, improving efficiency while ensuring equity and cost-effectiveness.

4 - Unveiling Operational Processes: the Impact of Real-Time Videos on Operational Outcomes

Zhanzhi Zheng, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Yuqian Xu, Bradley Staats

In the digital age, merchants are presented with an expanding range of resources to enhance operational transparency for customers. Moreover, the digital marketplace increasingly demands improved transparency due to the physical separation between merchants and customers. This paper explores a new method to enhance operational transparency for merchants on digital platforms: displaying real-time videos of their operational processes. Using a unique dataset about chain restaurants on a leading food delivery platform, we employ the difference-in-differences method with matching techniques for empirical analyses. Our results show that the display of real-time videos indeed improves restaurants' operational performance on the platform, including demand, customer average spending per order, and revenue. Furthermore, we find that reducing information asymmetry could be a key mechanism driving the positive effects of real-time videos. Specifically, these benefits are more pronounced for restaurants with less available information on the platform, such as those with fewer online reviews or lacking the premier label. Additionally, by leveraging advanced computer vision algorithms, we analyze the performance impact of key features in real-time videos through image mining, focusing on both the quality and content of the visual information that customers observe. Our findings indicate that both image quality and region of interest coverage play a crucial role in enhancing the benefits of operational transparency. Overall, our study highlights the importance of sharing visual information on operational processes to enhance performance in the digital marketplace, offering actionable insights for platform managers and service providers aiming to optimize their video content for improved outcomes.

TB06

Building A Level 3 A306

Platform Design

Invited Session

eBusiness

Chair: Kai Sun, University of Florida, Sunderland, MA, United States

1 - The Toll of Driver Moral Hazard in Ride-Hailing Platforms

Mengwei Qu, University of Connecticut, Storrs, CT, United States, Chen Liang, Chunxiao Li, Shuliu Yuan, Bin Gu

This paper investigates the prevalence of moral hazard behaviors among ride-hailing drivers and evaluates the impact of different forms of platform control on their engagement in such behaviors. By examining the differences in incentive structures across product lines within a prominent ride-hailing platform in China, we utilize rides where driver compensation is fixed as a benchmark to measure moral hazard in similar rides with variable compensation. Leveraging the quasi-random variation in trip characteristics arising from the platform's order dispatch algorithm, we find that (1) Ride-hailing drivers indeed engage in moral hazard behaviors, including taking detours and exceeding estimated payment. (2) During trips with higher compensation, e.g., long trips or those with a higher driver compensation ratio, drivers tend to exhibit increased, rather than decreased, engagement in moral hazard behaviors. (3) Drivers on the hook for a car lease contract exhibit significantly lower levels of moral hazard behavior than independent drivers.

2 - Policing the Platform: Uncovering the Impact of Police Involvement in Online Platforms on Users' Voluntary Contributions

Jingyan Dai, University of Florida, Gainesville, FL, United States, Qili Wang, Shu He, Liangfei Qiu

Government organizations are increasingly using social media to engage with the public, with police departments serving as a prominent example. However, the effects of government presence on online platforms, particularly on users' voluntary contributions, remain unclear. In this study, we address the gap and examine how police involvement influences users' voluntary contributions on a leading online neighborhood platform. Using a difference-in-differences approach, we find that police involvement leads to a decline in voluntary user posts. This reduction is primarily concentrated in posts concerning less severe and more ambiguous safety incidents, suggesting that users may engage in increased caution and self-censorship in response to perceived police monitoring. To better understand the underlying mechanisms, we conduct additional analyses by introducing moderators such as the percentage of visual evidence (e.g., photos and videos in posts) and the number of police-generated posts. Furthermore, we evaluate the effects of police involvement on the quality of user posts. Our findings show that the length and uniqueness of safety-related posts increase, indicating potential community safety benefits linked to police involvement. Overall, our study offers valuable insights for online platforms aiming to collaborate with government agencies in ways that enhance platform content quality while preserving user engagement. It also provides practical guidance for government agencies on leveraging social media to foster more effective community engagement.

3 - Multi-Homing Across Social Media: Influencer Behavior and Its Impact on Content Creation

Kinzhi Rao, University of Florida, Gainesville, FL, United States, Jingyan Dai, Yulin Hao, Liangfei Qiu

This study examines the strategies and impacts of multi-homing behavior among social media influencers. Using data from TikTok and RedNote, we provide empirical evidence on the current landscape of multi-homing: only a minority of influencers engage in multi-homing, typically doing so when follower growth on their original platform stagnates, and they choose to cross-post only a subset of their content to the other platform. To identify the causal effects of multi-homing on content creation and performance, we employ a Difference-in-Differences (DiD) approach. Our results show that multi-homing leads to significant increases in follower count, content quantity, user engagement, and sponsorship activity. Furthermore, multi-homing promotes content diversification and generates positive spillover effects on exclusive (non-multihomed) posts. These findings offer valuable insights into influencer strategies and platform competition, highlighting multi-homing as a powerful tool for audience expansion and monetization.

4 - Nudging for Re-Engaging Multi-Homing Consumers: Evidence from a Randomized Field Experiment

Xi Jing, Boston University, Boston, MA, United States, Tong Shen, Chen Liang, Chunxiao Li, Bin Gu

Multi-homing poses an increasing challenge for digital platforms, particularly in industries where aggregator apps enable users to submit multi-homing requests, allowing simultaneous product or service searches across multiple distinct platforms. To counteract multi-homing, platforms can deploy nudging strategies to encourage users to conduct product or service searches directly within the platform's native channel. To assess this strategy in a ride-hailing context, we leverage archival data from a randomized field experiment on a leading ride-hailing platform in China. The experiment involves 49,902 multi-homing users who had initiated at least one multi-homing request during the one-week experiment period. Among them, those randomly assigned to the treatment group received a message nudging them to use the focal platform's native channel upon their multi-homing request, while those in the control group received no intervention. We find interesting results. First, nudging significantly increases consumption on the platform through its native channel usage without cannibalizing rides booked via the aggregator. Second, nudging is more effective among users with a stronger brand attachment. However, it is not effective in bringing back disengaged multi-homing users to the platform's native channel. Third, the effect of nudging is contingent on cross-channel promotion comparisons. Nudging is only effective when the platform's native channel offers a larger discount than the aggregator. Fourth, the effect of nudging intensifies as the coupon validity period shortens. Our findings offer important practical insights for optimizing nudging strategies to re-engage multi-homing users and offer broader implications for platform competition dynamics.

TB07

Building A Level 3 A307

Routing and Scheduling

Job Market Showcase

Job Market Showcase

Chair: Amirhossein Moosavi, University of Michigan, Ann Arbor, United States

1 - A Model-Based Approach to Vacant Vehicle Routing of A Ride-Sourcing Fleet in Transportation Networks

Guocheng Jiang, University of Massachusetts Amherst, Amherst, MA, United States, Song Gao

A model-based approach to the optimal routing of vacant ride-sourcing vehicles is developed for scalability with respect to the fleet size in metropolitan scale transportation networks. The routing is formulated as an average-reward maximizing, infinite-horizon semi-Markov Decision Process problem in a physical transportation network with each node as a state and outgoing links as actions. All taxis are assumed to follow a uniform, parameterized, probabilistic policy to reduce the state and action space dimensions and balance out crowding. State transition probabilities are endogenous due to the competition among vehicles for passengers, and theoretical analyses are provided regarding the existence and uniqueness of the stationary distribution of the embedded Markov chain for a given policy.

A proof-of-concept two-node network case study shows that the optimal policy makes trade-off between moving to a high-demand area and the cost of relocation, while heuristics relying purely on demand distributions do not perform as well. A case study in a large network with around 30k links is conducted. The vacant taxi spatial distribution of the optimal policy indicates that vacant taxis tend to stay in and move towards areas with high demand such as the city center and transportation hubs, consistent with intuition. As the fleet size increases, the optimal policy dictates more widely spread vacant taxis and the optimal unit profit decreases, suggesting that competition leads to certain loss of profit.

2 - Distributed Load Balancing with Workload-Dependent Service Rates

Wenxin Zhang, Columbia University, New York, NY, United States, Santiago Balseiro, Robert Kleinberg, Vahab Mirrokni, Balasubramanian Sivan, Bartek Wydworski

We study distributed load balancing in bipartite queueing systems where frontends route jobs to heterogeneous backends with workload-dependent service rates. The system's connectivity---governed by compatibility constraints like data residency or resource requirements---is represented by an arbitrary bipartite graph. Each frontend operates independently without any communication with the other frontends, and the goal is to minimize the expectation of the average latency of all jobs. We propose a closed-loop policy called the Greatest Marginal Service Rate policy (GMSR) that achieves effective coordination without requiring arrival rates knowledge.

In a discrete-time stochastic model, we show that the behavior of our routing policy converges (almost surely) to the behavior of a fluid model, in the limit as job sizes tend to zero and job arrival rates are scaled so that the expected total volume of jobs arriving per unit time remains fixed. Then, in the fluid regime, we demonstrate that the policy attains an ϵ -suboptimal solution in $O(\delta \log(1/\epsilon))$ time from δ -suboptimal initial workloads, which implies global convergence to the centrally coordinated optimal routing.

Finally, we analyze the fluid model when the system is overloaded. We show that GMSR lexicographically maximizes throughput, maximizes the number of stable backends, and minimizes their collective workload.

3 - Rerouting During Shipping Lane Disruption

Yaniv Ravid, University of Toronto, Toronto, ON, Canada, Ming Hu

Major disruptions to global supply chains in recent years have imposed billions of dollars of additional costs onto consumers, retailers, shippers and manufacturers all around the world. In many instances, these disruptions reduce the capacity of major shipping lanes for unknown periods of time. These short-term disruptions, often driven by extreme weather, political forces, and chance events, have long-term implications on suppliers' decision-making. We study the implications of these disruptions on

shippers and the global supply chain using a transient queuing model. We derive the optimal self-interested routing policy under uncertain disruption length, and show that myopically dealing with this uncertainty results in unnecessarily long lines of ships awaiting travel in the disruption channel. We use real container-ship traffic data to show how our findings can mitigate the impact of such disruptions. We fit our model to the blockage of the Suez Canal in 2021, and analyze the sentiment of over 4,000 news articles to measure shippers' uncertainty in the disruption length, showing how hundreds of ships could have avoided unnecessary delays by choosing alternative routes, saving billions of dollars in delay costs imposed onto consumers as a result.

4 - Learning-Based Distributed Ambulatory Care Scheduling

Amirhossein Moosavi, University of Michigan, Ann Arbor, MI, United States

This work studies an ambulatory care scheduling problem for a geographically distributed healthcare center offering multi-appointment, multi-class, and multi-priority treatments. The problem is considered in a dynamic environment characterized by uncertain patient arrivals and emergency department utilization, formulated as an infinite-horizon Markov decision process. Given the limitations of conventional methods in solving large-scale instances, a neural network is integrated within the Markov decision process model to simplify feasibility constraints while ensuring adherence to the problem's assumptions. To mitigate the curse of dimensionality inherent in the Markov decision process formulation, an affine approximation architecture is employed to estimate the value function. Subsequently, an equivalent linear programming model is developed and solved using column generation, enabling computation of approximate optimal policies. Additionally, two straightforward, easy-to-implement scheduling policies are derived from this approach. Simulation results indicate that the approximate optimal policy and derived heuristic rules significantly outperform alternative scheduling methods. Specifically, the proposed approach enhances performance metrics within a large Canadian ambulatory care center context. In contrast, the traditional template-based scheduling rule currently used at the center results in high resource utilization but often leads to suboptimal scheduling decisions. Providing booking clerks with efficient, data-driven scheduling rules demonstrates substantial improvements over scheduling templates.

TB08

Building A Level 3 A310

Generative AI in Business Analytics: Innovations, Applications, and Implications

Invited Session

Information Systems

Chair: Houping Xiao, Robinson College of Business, GSU, Atlanta, GA, United States

1 - Rewiring Deep Engagement: How Generative AI Transforms Knowledge Exchange in Online Q&a Communities

Xinyuan Wei, Georgia State University, Atlanta, GA, United States, Xinfu Fu, Arun Rai, Yusen Xia

This research examines how the emergence of generative AI (GenAI), specifically ChatGPT, impacts user engagement within online Question & Answer (Q&A) communities like Stack Exchange. Drawing on a collective information foraging perspective, the study analyzes user behavior before and after ChatGPT's release, treating it as an exogenous technological shock. Utilizing a difference-in-differences approach, the research focuses on shifts in answer elaboration and evaluation, key aspects of collective knowledge foraging. Findings indicate that GenAI increases deep engagement by broadening both answer elaboration and total evaluation breadth, partially mediated by changes in question evaluation. This suggests GenAI not only serves as a substitute for user-generated content but also stimulates wider distribution of user attention and engagement across diverse answers, enhancing the collective foraging process. The study introduces the concept of cognitive discrepancy between AI- and peer-generated content as a driver for deeper engagement in this foraging ecosystem and offers insights for platform design in an AI-mediated environment, adapting to evolving social information foraging dynamics.

2 - A Self-Instantiated Multi-Agent System for Simulating Field Experiments

Yaoyu Chen, University of Illinois Chicago, Chicago, IL, United States, Yingda Lu, Yuheng Hu

Field experiments are a gold standard in social science research for establishing causal claims in naturalistic settings. However, they are notoriously costly, time-consuming, ethically sensitive, and difficult to replicate. To address these challenges, this paper proposes a novel framework that uses Large Language Models (LLMs) to simulate field experiments through a Self-Instantiated Multi-Agent System (SIMAS). LLM-driven agentic models are well-suited for simulating individual decision-making, as they instantiate autonomous entities with distinct goals and internal states. Building on this capacity, recent studies have demonstrated the feasibility of replicating laboratory experiments using LLMs. Extending this line of research, our work targets the more complex and underexplored domain of field experimentation by offering a fine-grained simulation of naturalistic decision-making processes. In contrast to prior lab-based simulations—often constrained by limited generalizability and lacking coordinated multi-agent dynamics—the proposed SIMAS is adaptive to universal field experiment scenarios, capable of determining treatments, automatically generating synthetic participants, collecting responses, and drawing experimental conclusions through a reflective, iterative process. Notably, its self-instantiation mechanism enables the large-scale generation of synthetic agents, fulfilling the scale requirements of various kinds of experiments. Our artifact is to be evaluated against a diverse set of field experiments across domains such as finance, marketing, and organizational behavior. This study contributes both a theoretical advancement—demonstrating emergent capabilities of LLMs in simulating human-like interactions at scale—and a methodological innovation in designing domain-adaptive multi-agent architectures for experimental research.

3 - Evaluating Generative AI for Medical Reporting: a Comparative Study of Clinician and LLM-Based Diagnoses for Mild Cognitive Impairment

Houping Xiao, Robinson College of Business/GSU, Atlanta, GA, United States

Generative Artificial Intelligence (AI), such as ChatGPT, holds significant promise in healthcare, particularly in automating the generation of medical reports and clinical recommendations. This capability raises important questions about its potential to complement or even replace traditional diagnostic processes, such as those used for Alzheimer's Disease (AD). Clinician-generated medical reports based on functional MRI (fMRI) images have long served as a gold standard in cognitive assessments. However, it remains unclear how patients and healthcare professionals perceive and evaluate reports produced by large language models (LLMs) compared to those authored by clinicians. This study investigates the effectiveness of LLM-generated medical reports in the context of predicting Mild Cognitive Impairment (MCI). By comparing medical reports written by clinicians and LLMs, and incorporating analysis of impaired shallow spatial features derived from deep learning models, we aim to assess the diagnostic reliability and user appraisal of AI-generated content. The findings will offer critical insights into the role of generative AI in clinical decision-making and its broader implications for the future of healthcare delivery.

4 - An Integrative ML and Rag Pipeline for Health Data-Based Hypothesis Generation

mouayed lajnef, Georgia State University, Atlanta, GA, United States, Aaron Baird, Yusen Xia, Arun Rai

Hypothesis generation in scientific research has traditionally relied on researcher intuition, gap spotting, and theory-driven inquiry. Recent advances in machine learning (ML) and large language models (LLMs) offer new opportunities for data-driven discovery by surfacing complex patterns in observational data. However, two major challenges persist: (1) ML-generated insights often lack interpretability and standardization, and (2) verifying the novelty of these hypotheses typically requires extensive manual literature reviews, which are time-consuming and difficult to scale. To address these issues, we propose an integrated hypothesis generation and verification pipeline that combines interpretable ML methods, Retrieval-Augmented Generation (RAG), and a structured conceptual framework. Our pipeline includes a five-step ML discovery module that uncovers statistically robust higher-order interactions, followed by a RAG-powered literature agent that retrieves and synthesizes relevant studies from PubMed to assess novelty. To better align machine-generated hypotheses with existing scientific knowledge, we extend the widely-used PICO framework (Population, Intervention, Comparison, Outcome) into PICO-CMV by introducing three additional dimensions: *Conditions*, *Modifiers*, and *Validation*. This enhancement allows for a more granular organization of evidence and supports structured comparison with existing literature. A Human-in-the-loop design is implemented at multiple stages, including hypothesis refinement, model interpretation, and literature verification, ensuring that domain expertise shapes and validates machine-driven insights. We illustrate the utility of this pipeline through an application in the substance use disorder (SUD) context, showing how novel, testable hypotheses can be efficiently surfaced and contextualized within the existing body of evidence.

TB09

Building A Level 3 A311

Strategy and Innovation

Job Market Showcase

Job Market Showcase

Chair: Sidi Zhao, Auburn University, 1761 Wire Rd, Apt 6-11, Auburn, 36832, United States

1 - Incoherence of Business Model Innovation During Digital Transformation**Chen Ye, University of Cambridge, CAMBRIDGE, United Kingdom**

This research investigates business model incoherence during digital transformation, a critical challenge for innovating firms. Initial foundational work, based on 50 interviews and workshops, classified six types of incoherence—strategic, organizational, operational, partner-driven, market-driven, and regulation-driven—stemming from four primary sources: changes in decision rights, incentives, information flow, and material flow, alongside potential solutions. Furthermore, focused longitudinal single-case studies with firms examined the mechanisms of incoherence through the paradox lens, highlighting how firms navigate contradictory demands like efficiency versus flexibility by managing different business models. Building on this, multi-case studies with the firms further explored the deep mechanisms of how incoherence occurs and is managed within different value configuration models. This demonstrated how "disciplined incoherence"—loose coupling and organized contestation—is used to manage these tensions, revealing alternative effects on creating versus sustaining competitive advantage. Additionally, industry-specific multi-case studies also explored incoherence mechanisms, such as in Large Language Model ecosystems (LLM), where firms manage strategic tensions, and in predictive maintenance in manufacturing, where institutional constraints drive incoherence during servitization.

Moreover, the study expanded by using system dynamics modeling on firms to reveal how misalignments initiate incoherence, identifying balancing feedback loops and reinforcing loops in contexts. Additionally, a crisp-set qualitative comparative analysis (csQCA) of cases identified specific configurations of sources leading to internal incoherence versus external incoherence. Collectively, these studies deepen the understanding of incoherence, extending strategic contradiction theory and the Business Model Cohesiveness Scorecard (BMCS) by detailing dynamic pathways and offering actionable solutions for managers to anticipate and resolve the incoherence.

2 - Incorporating Sustainability into Portfolio Choices: a Rational Investor Approach**Giulia Crippa, Princeton University, Princeton, NJ, United States**

Traditional mean-variance portfolio frameworks often impose constraints—such as minimum sustainability thresholds—as proxies for investor preferences, implicitly assuming that constrained and preference-based approaches yield equivalent solutions. However, this assumption may misrepresent rational investor behavior or fail to capture investor heterogeneity.

In this paper, we extend the classical mean-variance framework by incorporating a sustainability attribute directly into the investor's utility function. We formally characterize the conditions under which constrained and preference-based formulations are equivalent, building on prior results from the dynamic portfolio choice literature. In a complete market setting, we derive closed-form equilibrium strategies that generalize the classical solution to account for sustainability preferences.

We validate our theoretical findings through simulations and an empirical application using U.S. equity data. By endogenously recovering investor trade-offs across return, risk, and sustainability, we construct a three-dimensional mean-variance-sustainability (MVS) efficient frontier that yields Pareto-optimal portfolios. Our framework establishes a coherent and flexible foundation for sustainable investment design, with direct implications for asset managers and policymakers seeking to align portfolio performance with sustainability objectives.

3 - Contingent Performance Impact of Digital Transformation Orientation: An Organizational Change Perspective**Hongyu DUAN, Hong Kong Baptist University, Hong Kong, Hong Kong**

Despite growing global emphasis on the potential of digital transformation as a strategic orientation, firms often struggle to realize its anticipated benefits. To clarify the intricate relationship between digital transformation orientation (DTO) and firm performance, this study proposes that organizational changes exert powerful influences. A longitudinal data set pertaining to 2,766 Chinese public firms, for the years 2010–2022, reveals that on average, DTO diminishes firm performance (return on assets), but this negative impact also is moderated by specific organizational changes, both asset-centric (e.g., asset divestiture, expansion) and human-centric (e.g., executive and employee turnover). Asset divestitures and expansion mitigate the negative impacts of DTO by optimizing resource allocations; employee turnover introduces new skills that can alleviate these adverse performance impacts. However, frequent IT executive turnover exacerbates the negative performance impacts by disrupting strategic persistence and leadership stability. The combined results thus help address prior mixed findings regarding the performance implications of DTO and highlight contingencies stemming from organizational changes. In underscoring the importance of aligning organizational change initiatives with DTO, the current study offers actionable recommendations for firms navigating their own digital transformation journeys.

4 - How Developers' Social Interaction on External Platforms Shapes Firms' Strategy Orientations in Digital Innovation**Sidi Zhao, Auburn University, Auburn, AL, United States, Pei Xu, Uzma Raja**

Digital innovation is essential for firms aiming to enhance competitiveness, improve operations, and drive value creation. A firm's strategic orientation toward digital innovation shapes how it engages with emerging technologies and determines its long-term success. While existing research emphasizes top management as the primary driver of this orientation, the role of developers, especially through external social interactions, remains underexplored. Drawing on the absorptive capacity theory and heterogeneous diffusion model, we propose that developers play a critical role in steering their firms' digital innovation strategies. Collaborative development platforms with social media features serve as key conduits for knowledge transfer beyond firm boundaries, allowing developers to absorb external expertise and inform strategic direction from the bottom up. Leveraging longitudinal data from 2010 to 2022 on USPTO patents and GitHub developer networks, we identify four distinct innovation strategies firms may pursue: No Engagement, Comprehensiveness-Only, Efficiency-Only, and Fully-Engaged. We then examine how developer network characteristics, social proximity (cohesion and role equivalence), infectiousness, and susceptibility, shape firms' strategic choices. Our results show that while social cohesion and susceptibility have mixed effects, role equivalence and infectiousness consistently exert positive influences. Moreover, susceptibility moderates these effects, dampening the influence of cohesion but strengthening that of role equivalence. This study contributes a novel typology of digital innovation strategies and advances a developer-driven view of innovation. By highlighting how micro-level developer interactions drive macro-level strategic outcomes, we complement traditional top-down accounts and offer new insights into how firms can leverage developer communities for sustained innovation and competitive advantage.

TB10

Building A Level 3 A312

Novel Models in Game Theory and Social Sciences

Job Market Showcase

Job Market Showcase

Chair: Ali Akhavan, MIT, Boston, MA, United States

1 - Efficiently Solving Variational Inequalities Under the Minty Condition, and Game-Theoretic Applications**Brian Zhang, Carnegie Mellon University, Pittsburgh, PA, United States**

Solving (Stampacchia) variational inequalities (VIs) is a foundational problem at the heart of optimization, with a host of critical applications ranging from engineering to economics. However, this expressivity comes at the cost of computational hardness. As a result, most research has focused on carving out specific subclasses that elude those intractability barriers. A classical property that goes back to the 1960s is the Minty condition. We will establish the first $\text{poly}(d, \log(1/\epsilon))$ -time algorithm for solving SVI problems under the Minty condition. Prior approaches either incurred an exponentially worse dependence on (and other natural parameters of the problem) or made overly restrictive assumptions – such as strong monotonicity. To do so, we introduce a new variant of the ellipsoid algorithm wherein separating hyperplanes are obtained after taking a gradient descent step from the center of the ellipsoid. It succeeds even though the set of SVIs can be nonconvex and not fully dimensional. Moreover, when our algorithm is applied to an instance with no MVI solution and fails to identify an SVI solution, it produces a succinct certificate of MVI infeasibility. Finally, we will discuss the numerous relationships and implications of our framework for learning in games and computational game theory more broadly, including the best algorithms for computing and learning certain kinds of correlated equilibrium in concave games.

This talk will cover joint works with Ioannis Anagnostides, Emanuel Tewolde, Ratip Emin Berker, Gabriele Farina, Vincent Conitzer, and Tuomas Sandholm.

2 - Online Learning for Dynamic Vickrey-Clarke-Groves Mechanism in Sequential Auctions Under Unknown Environments

Vincent Leon, University of Illinois Urbana-Champaign, Urbana, IL, United States, Rasoul Etesami

We consider the problem of online mechanism design for sequential auctions in unknown environments, where the underlying market and the values vary over time as the auctions progress. We model the sequential auctions as an infinite-horizon Markov decision process (MDP), where the transition kernel and reward functions are unknown to the seller. In each round, the seller determines an allocation and a payment for each bidder. Each bidder receives a private reward and submits a sealed bid to the seller. The state, which represents the underlying market, evolves according to an unknown transition kernel and the seller's allocation. Unlike existing works that formulate the problem as a multi-armed bandit model or as an episodic MDP, where the environment resets to an initial state after each episode, our paper considers a more realistic and sophisticated setting in which the market continues to evolve without restarting. We first extend the Vickrey-Clarke-Groves (VCG) mechanism, which is known to be efficient, truthful, and individually rational for static auctions, to sequential auctions, thereby obtaining a dynamic VCG mechanism counterpart that preserves these desired properties. We then focus on the online setting and develop a reinforcement learning algorithm for the seller to learn the underlying MDP model and implement a mechanism that closely resembles the dynamic VCG mechanism. We show that the learned online mechanism asymptotically converges to a dynamic mechanism that approximately satisfies efficiency, truthfulness, and individual rationality with arbitrarily high probability and achieves guaranteed performance in terms of various notions of regret.

3 - An I-Mobility" Framework for Studying Social Mobility: Individualized Inference via Generative Analysis of Discrete Data

Jiawei Huang, University of Cincinnati, Cincinnati, OH, United States, Dungang Liu, Yuan Jiang

Social mobility refers to the ability of individuals or groups to move within a social hierarchy. Interest in social mobility has grown over the past decades due to rising concerns over educational disparities and the intergenerational persistence of poverty. Existing literature primarily investigates this issue by focusing on broad demographic groups (e.g., race, gender, or country), which may overlook important nuances between individuals. In this work, we establish an 'i-Mobility' framework that allows us to study social mobility at the individual level. Specifically, given a predefined profile with a combination of individual characteristics, our framework can provide a measure that reflects the mobility of this 'individual.' Analysis of the well-regarded General Social Survey (GSS) shows that our framework is more robust than traditional group-focused methods for studying social mobility. Moreover, our framework can capture heterogeneity at the individual level and distinguish between different profiles by considering more nuanced personal characteristics.

4 - Benchmarking Estimation of Dynamic Models in Social and Behavioral Sciences

Ali Akhavan, MIT & MGH, Cambridge, MA, United States, Jimmy Cabot, Hazhir Rahmandad

Reliable parameter estimation is critical for dynamic simulation models across various fields, such as economics, epidemiology, ecology, marketing, management, and sustainability. We present a benchmarking framework that systematically assesses the inference quality of diverse estimation methods for such models. The framework provides various performance metrics, such as RMSE, posterior contraction, interval score, z-score, and simulation-based calibration tests. In addition, we offer a testbed of canonical dynamic models with different scales (e.g., number of parameters) and dynamic complexity levels (e.g., different modes of behavior). The testbed includes representative models such as Bass Diffusion, SEIRb, Market Growth, and Lotka-Volterra, selected through a literature review and a survey of modeling experts. We use the benchmarking framework to assess Neural Posterior Estimation (NPE) as an illustrative example of emerging likelihood-free approaches and report the inference quality as a function of various training data sizes. The method-agnostic testbed models and performance metrics will allow researchers to assess and compare alternative estimation methods, ranging from classical maximum-likelihood estimators to various simulation-based inference algorithms such as Neural Ratio Estimation (NRE) or Sequential Neural Posterior Estimation (SNPE). The benchmark can also be expanded with new testbed models that may incorporate more complex estimation problems. We hope our research promotes a transparent, scalable, and replicable way of quantitative evaluation of inference quality across the existing and emerging estimation methods for dynamic simulation models.

TB11

Building A Level 3 A313

AI, Computer Vision, and Virtual Reality

Job Market Showcase

Job Market Showcase

Chair: Nahian Tahmin, The University of Tennessee Knoxville, Knoxville, TN, United States

1 - Metadata-Aware Multimodal Object Detection with Feature-Level Fusion for Unpaired Rgb-Thermal Datasets

Nicolas Bustos, University of South Florida, Tampa, FL, United States, Susana Lai-Yuen, Tapas Das

Unpaired multimodal datasets pose challenges to fusing information in multimodal object detection algorithms. Unpaired datasets either lack spatial alignment between image modalities or have missing images for some scenes. Traditional solutions rely on image registration (e.g., SIFT, Fourier-Mellin Transform) and domain translation methods (e.g., CycleGAN). However, these techniques are often sensitive to environmental conditions—especially when using RGB images—and are not optimized for detection tasks. Instead, they produce pseudo-realistic aligned pairs that are not often suitable representations to enhance object detection. Moreover, metadata such as weather or time-of-day labels remains underutilized despite its potential to improve object detection.

We propose a novel multimodal object detection framework capable of leveraging all image modalities from datasets that are unpaired. A thermal-to-RGB generator is trained to produce synthetic RGB features optimized for object detection, which uses a YOLOv10 model pre-trained on RGB images. These RGB features are fused in another YOLOv10 model during its training with original thermal images. At this step, we also fuse metadata features extracted from ResNet50 models pre-trained to classify weather and time of the day on the thermal images.

The Teledyne FLIR dataset is used to train and test the models in our framework. This dataset contains unpaired RGB and thermal images, partially labeled with weather and time of day. We train the algorithms with only annotations for cars and people, as these are the most critical categories in the fields where multimodal object detection is applied, such as surveillance and autonomous driving.

2 - Human-Centric Xr Design to Support Dimensions of Wellness: Applications Across Domains and Age Groups

Emre Eraslan, University of Illinois Urbana-Champaign, Urbana, IL, United States, Qiyuan Cheng, George Mois, Wendy Rogers, Avinash Gupta

Extended reality (XR), which includes virtual reality (VR), augmented reality (AR), and mixed reality (MR), presents new opportunities for enhancing human well-being across different life stages. Although XR technologies are increasingly applied in health, education, and social domains, there is limited understanding of how specific XR modalities correspond to the diverse and interconnected Dimensions of Wellness (DoW), especially when considering differences among young, middle-aged, and older adults. The DoW framework includes emotional, physical, intellectual, social, occupational, environmental, financial, and spiritual aspects, each contributing to overall quality of life. This study offers a systematic mapping review that classifies XR interventions by DoW, age group, and study context. Following the PRISMA-ScR methodology, 145 peer-reviewed studies published between 2016 and 2024 were identified through searches in PubMed, Scopus, and Web of Science. The analysis shows that fully immersive VR is the most frequently used modality, while non-immersive VR, AR, and MR show more variable patterns of use. Emotional, physical, intellectual, and social DoWs are consistently addressed across age groups, and occupational DoW appears primarily in studies involving young adults. In contrast, environmental, financial, and spiritual DoWs remain largely unaddressed. These findings highlight critical gaps in current XR research and suggest opportunities for more inclusive, age-responsive design. By mapping DoW priorities to XR capabilities, this study paves the way for designing targeted VR applications that address population-specific needs, bridging the gap between research insights and real-world interventions.

3 - Optimization and Learning Methods Applied to Automated Vision-Based Body-in-White Dimensional Inspections

Yifei Zhang, Ohio State University, UPPER ARLINGTON, OH, United States, Theodore Allen

Traditional measurement of gaps and flush in automotive Body-in-White (BIW) assembly relies on manual instruments such as feeler gauges or dial gauges, which are labor-intensive, time-consuming, and unable to provide continuous data. This work develops a vision-based inspection system capable of automatically and non-invasively measuring BIW dimensions under dynamic production conditions.

Due to challenges such as limited supervised dimensional data, scarce target detection images, moving targets, and the requirement for prediction errors below 0.5 mm, this work proposes a framework that integrates Optimal Transport (OT) theory, Markov Decision Process (MDP) formulations, and regression-based modeling. OT-based image matching with Procrustes Analysis ensures precise alignment of edge points, enabling robust geometric transformation learning. The MDP formulation allows Reinforcement Learning (RL) agents to autonomously learn optimal tracking strategies in dynamic environments, while polynomial regression provides accurate dimensional prediction with limited supervision.

Experimental results show that the system achieves over 93% detection accuracy and reduces computation time by more than two orders of magnitude compared with grid search baselines. These findings demonstrate that combining OT-based matching, RL-driven tracking, and regression modeling yields a scalable, low-cost solution for high-precision visual inspection.

4 - Advancing Admixture Genetic Analysis with Large Language Model (AdmixAI)

Nahian Tahmin, The University of Tennessee Knoxville, Knoxville, TN, United States, Lokesh Chinthala, Franco Marsico, Akram Mohammed, Silvia Buonaiuto, Vincenza Colonna, Tesfaye Mersha, Robert L Davis, Anahita Khojandi

Recent advancements in artificial intelligence have significantly enhanced genomic research. LLMs facilitate the integration of contextual information into predictive models, thereby improving our understanding of ancestry-informed genetic influences on disease predictions.

Despite these advancements, genomic studies have predominantly focused on populations of European ancestry, leading to a limited understanding of complex diseases in admixed populations, such as self-reported African-American (AA) individuals. This leads to limitations of the generalizability of risk prediction models, exacerbating disparities in disease diagnosis and treatment. Additionally, most previous studies have primarily relied on global ancestry estimates and linear associations between genetic variation and ancestry. This approach often overlooks the complex, nonlinear interactions that can occur at more granular levels, such as local ancestry (LA). Incorporating local ancestry (LA) can offer deeper insight into genetic disease risk, especially in admixed individuals. Notably, conditions like asthma exhibit significant prevalence variations across diverse populations, underscoring the necessity for ancestry-informed methodologies in genomic risk assessment.

To address these challenges, we propose AdmixAI, an innovative LLM-based pipeline that integrates genotype data, LA ancestry data and clinical data to enhance disease risk prediction in self-reported AA populations. AdmixAI explicitly models interactions between ancestry and genetic variants to enhance predictive accuracy and support precision medicine in admixed populations. To our knowledge, it is among the first applications of LLMs for ancestry-informed genomic risk prediction, specifically targeting asthma. By integrating ancestry and genotype data into LLM-based pipelines, AdmixAI aims to refine polygenic risk scoring and promote more equitable precision genomics.

TB12

Building A Level 3 A314

Applications of Vehicle Routing in Modern Logistics

Contributed Session

Contributed

Chair: Nikhil Eti, Clemson University

1 - A Practical Algorithm for Multi-Vehicle Residential Waste Collection Problem

Seungyeop Lee, Pohang University of Science and Technology (POSTECH), Pohang, Korea, Republic of, Sangil Han, Byung-In Kim

The residential waste collection problem is a variant of the arc routing problem. It seeks an arc sequence that minimizes the total route cost of a fleet of heterogeneous vehicles servicing all required arcs and edges on a mixed graph, which includes both directed arcs and undirected edges. Each vehicle must visit landfills multiple times during its route to dump collected waste due to capacity constraints. It must also comply with a maximum work time from departure to return at the depot. However, solutions that focus solely on minimizing total route cost are often unacceptable to practitioners for the following reasons. First, they neglect turn penalties associated with heavy vehicles, leading to inaccurate calculations of work time. Second, unbalanced workloads among drivers are problematic due to disparities in work intensity. Third, excessive overlap in service areas and repeated visits to the same roads are undesirable for both drivers and residents. To address these issues, we propose a graph partitioning-based matheuristic. Specifically, we introduce a clustering algorithm that leverages both the graph structure and coordinate data of the road network to generate visually attractive clusters. Using this algorithm, we adopt a sequential clustering approach, where balanced clustering is first used to assign required arcs and edges to each vehicle. Then, capacitated clustering is applied within each vehicle's cluster to assign the required arcs and edges to feasible subtrips. For each decomposed single-trip problem, we minimize turn-penalized route cost. Consequently, our method yields solutions that are cost-efficient, workload-balanced, and visually attractive.

2 - Joint Optimization of Driver Shift Assignment and Vehicle Routing in Ada Paratransit

Shijie Chen, Florida State University, Tallahassee, FL, United States, Hishamur Rahman, Nikola Markovic, Yanshuo Sun

Paratransit services mandated by the Americans with Disabilities Act (ADA) play a vital role in providing accessible transportation for individuals with disabilities. Traditional scheduling methods, however, often fail to account for drivers' shift preferences and the financial impacts of shift reassignments. To address this gap, we develop an integrated optimization framework that jointly determines vehicle routes and driver shift assignments for ADA paratransit. The problem is formulated as a route-based integer program, which requires a large number of routes and causes an infeasibility issue. Therefore, we propose a two-phase relaxation-decomposition algorithm that combines Lagrangian relaxation with column generation to address the challenges. Comprehensive numerical experiments using synthetic and real-world data reveal that our model achieves up to 42.8% cost reduction by optimizing shift reassignments and reducing reliance on extra board drivers. Additionally, the model adapts effectively to various operational scenarios, including cases with skewed shift preferences and rising shift differential costs.

3 - Last-Mile Delivery Assignment Across a Traditional Route and a Crowdsourcing Platforms

Raghavan Srinivasan, Metro State University, St.Paul, MN, United States, Joseph Szmerkovsky

In this study we aim to determine the assignment of packages to be delivered between traditional routes and crowdsourcing services using a contemporary Machine Learning (ML) approach. The objective is to achieve the lowest possible operational costs using a combination of the two delivery options given capacity constraints. To achieve this, a ML model is first used to determine the packages to be assigned to a crowdsourcing platform with the remainder being delivered using a traditional delivery route. The packages are to be delivered in a definite geographic region, and a Geographic Information System (GIS) is used to determine the optimal traditional route.

4 - Incentive-Based Consolidation In Last Mile Parcel Delivery

Semih Boz, University of Massachusetts Amherst, Amherst, MA, United States, Ahmed Ghoniem, Amro El-Adle

We study incentive-based parcel consolidation within the context of last-mile delivery by a single vehicle over multiple time periods. We propose a mixed-integer program enhanced with RLT cuts. The efficacy of our methodology is assessed through computational experiments on realistic networks.

5 - The Vehicle Routing Problem with Transfers

Maichel M. Aguayo Bustos, Universidad San Sebastián, Concepcion, Chile

In the vehicle routing problem (VRP), a fleet of capacitated vehicles serves a set of customers with known demands at minimum cost. We study the VRP with transfers (VRP-T), where customers can be visited by multiple vehicles. Additionally, customer locations can be used as transfer points to exchange loads among vehicles. We find that routing costs can potentially be reduced by up to 50% by allowing such transfers in the VRP. We develop a two-index mixed-integer programming (MIP) model and a heuristic algorithm based on multi-start local search (MSLS). The solutions obtained by MSLS are used as a warm start for the MIP, and the corresponding approach is tested. Computational results indicate substantial savings in costs incurred and a reduction in the number of vehicles required when transfers are allowed.

6 - Multi Depot Multi Trip Vehicle Routing Problem with TIME Windows

Daniel Neira, Virginia Tech, Blacksburg, VA, United States

The multi-depot multi-trip vehicle routing problem with time windows (MDMTVRPTW) is a complex optimization problem that extends traditional vehicle routing problems by incorporating multiple depots, multiple trips per vehicle, service-dependent loading times, and fixed destinations. The objective is to minimize the total distance traveled while ensuring customer visits within specified time windows and operational constraints.

This study presents new mathematical formulations for the MDMTVRPTW, including a compact two-index formulation based on depot node insertion and a three-index formulation using arc replenishment. These models efficiently handle vehicle capacity, trip duration limits, and service times. Additionally, we propose a heuristic method based on a secondary graph in a dynamic programming algorithm to improve computational efficiency. The heuristic is compared against the exact formulations to evaluate solution quality and computational performance.

We validate our approaches using well-known VRP benchmark instances, including Solomon's datasets and others, to ensure robustness and applicability. The results demonstrate that the heuristic achieves near-optimal solutions with significantly reduced computational time, making it suitable for large-scale instances. The comparison highlights the trade-off between exact optimization and heuristic efficiency.

Our findings contribute to vehicle routing research by introducing novel formulations and a practical heuristic for solving real-world logistics challenges. Future work will explore hybrid metaheuristic techniques to further enhance solution quality and scalability.

7 - Online Mapping Errors and Their Impact On Last-Mile Delivery Operations

YUQING CHENG, Arizona State University, Tempe, AZ, United States, Elliot Rabinovich, Stanley Lim

We investigate the magnitude of online mapping errors and their impact on last-mile delivery operations in urban environments. These errors are reflected in the distances that exist between the actual geographical coordinates for delivery addresses and the coordinates for these addresses in the online navigation maps used by last-mile delivery agents. We report on the magnitude of these errors. Moreover, we estimate these errors' effects on delays incurred by agents performing these deliveries as well as on the likelihood of missed deliveries. We close with policy implications for last-mile delivery services based on these findings.

8 - Multi-Stage Stochastic Multi-Door Cross-Dock Scheduling Under Uncertain Arrivals

Nikhil Eti, Clemson University, Greenville, SC, United States, Bill Ferrell, Nathan Huynh

Cross-docks are facilities in a logistics network that are used to consolidate shipments so that truck utilization is improved. Products requiring delivery to different destinations arrive at a cross-dock from multiple sources. They are then consolidated by destination and shipped on trucks that are full or nearly full. This has many benefits including lower costs and fewer trucks required for deliveries which reduces road congestion and emissions. This research focuses on cross-dock operations under disruptions; specifically, disruptions that delay arrival times of inbound trucks. This is important because outbound trucks have departure deadlines designed for on-time delivery to customers. Delays to inbound freight can delay the outbound truck awaiting these items but they can also cause a dock door to be occupied that is needed to load subsequent outbound trucks. A Multi-Stage Stochastic Programming (CD-MSSP) model of cross-dock operations has been developed to schedule outbound trucks so that delays to outbound departures are minimized. The model is used to solve a number of scenarios which involve disruptions to inbound trucks causing them to arrive late at the cross-dock. To solve the resulting large mixed integer models, a Progressive Hedging (PH) algorithm is used to find good feasible solutions. The PH algorithm relaxes the complicating constraints in CD-MSSP to enable scenario-based decomposition which promotes faster solution times. The PH algorithm is compared with direct branch and bound algorithm for various numerical test instances, and the PH algorithm is found to be up to 42 times faster.

TB13

Building A Level 3 A315

Inside Orange Toolbox: Data Science in Action at The Home Depot

Panel Session

Career Fair

Co-Chair: Lauren Maire, Home Depot, Atlanta, GA, United States

1 - Moderator Panelist

Lauren Maire, Home Depot, Atlanta, GA, United States

In this presentation we will discuss three exciting data science projects in progress at the Home Depot:

- The application of a deep learning AI approach to estimating price impact on demand and an approach to the problem of divergent scales in time series modeling
- How Home Depot leverages language models to deliver superior search relevancy, query understanding, and innovative natural language filtering enabling them to move beyond traditional token and facet-based mechanisms

Explore the development of a conversational BigQuery agent that is integrated into a Slack App and translates simple user questions into complex, executable SQL queries.

2 - Panelist

Anvar Hubs, The Home Depot, Atlanta, GA, United States

3 - Panelist**Harish Nair, The Home Depot, Atlanta, GA, United States****4 - Panelist****Jared Babcock, The Home Depot, Atlanta, GA, United States****5 - Panelist****Kevin Lee, The Home Depot, Atlanta, GA, United States****TB14**

Building A Level 3 A316

Behavioral Topics in Human-Algorithm Connection

Invited Session

MSOM: Technology, Innovation, and Entrepreneurship

Chair: Ruth Beer, Baruch College, CUNY, New York, NY, United States

Co-Chair: Evgeny Kagan, Johns Hopkins Carey Business School, Baltimore, MD, United States

1 - Backwards Planning with Generative AI**Samantha Keppler, University of Michigan, Ann Arbor, MI, United States, Park Sinchaisri, Clare Snyder**

Generative AI technologies can help workers do tasks faster and better. However, workers often must plan their tasks in addition to completing them, and this planning can be difficult. K12 teachers plan tasks (i.e., activities, quizzes, projects) by working backwards from end-of-year goals defined by state standards. In this paper, we ask: How are teachers, who must both plan and do tasks, beginning to use generative AI? We conduct a longitudinal case study of 24 US public school teachers. During the 2023--2024 school year, we gather from these teachers 360 minutes of recorded observation of generative AI use for their own work involving over 200 inputted prompts and responses, together with 29 in-depth interviews and 34 generative AI use surveys. Analyzing this data corpus, we find teachers separate into three groups: (1) those who seek generative AI input (i.e., thoughts or ideas about learning plans) and output (i.e., quizzes, worksheets), (2) those who only seek generative AI outputs, and (3) those not using generative AI. The teachers in the first group—but not the second group—report productivity gains in terms of workload and work quality. In sum, workers can use generative AI for known tasks (task-level) or for deciding which tasks to do (workflow-level). Task-level use may save time completing tasks and increase the number of tasks done, but workflow-level use can save time deliberating about which tasks to do and nudge people toward the "right" tasks.

2 - Algorithm Reliance and the Value of Information in Recommendation Systems**Ruth Beer, Baruch College, CUNY, New York, NY, United States, Daniela Saban**

We study individuals' preferences for delegating managerial decisions to a matching algorithm, after a large set of alternatives have been screened and the best candidates have been identified. Two behavioral considerations are central to our setting: algorithm reliance (i.e., individuals' willingness to delegate decision rights) and preferences for information (i.e., individuals' preferences for making a more informed decision and learning). Our experimental results show that most subjects significantly overpay for the option to gather more information and select the candidate themselves. In addition, their beliefs about the probability that the algorithm selects the worst candidate, and the expected difference in benefit between the candidates conditional on the algorithm selecting the worst candidate, are significantly higher than what the normative theory prescribes, suggesting that cognitive biases are a salient driver of behavior.

3 - Privacy, Risk and Fatalism**Cagla Ergul, University of Michigan, Ann Arbor, MI, United States, Stephen Leider, Ruslan Momot**

Firms collect and analyze users' data to personalize and improve their service. However, the risk of data breaches exposing a user's personal data imposes privacy costs on the user. In this paper, we experimentally study how users engage with the firms' services under different levels of data breach risk. To motivate our experiment, we first develop a theoretical model to characterize how the users' behavior changes as the background breach risk from previous service use increases. We find that privacy-conscious users decrease their service usage as the firm collects more data when the background breach risk is higher. However, if the users have fatalistic beliefs (i.e. they believe they will eventually face a breach, and become less sensitive to further breach risks), they do not protect their privacy as the background breach risk increases. We conduct a laboratory experiment, examining both the low and high background breach risk cases. We find that the majority of the users are fatalistic and are unresponsive to the Firm's increasing data collection. Users primarily choose the action that maximizes payoffs conditional on a breach happening. The share of fatalistic users is similar for both levels of background breach risk. Our results align with the fatalism predictions of the model; users are not sensitive to further data collection regardless of the background risk. Our results still hold when breach risk is communicated in a disaggregated format: most users are fatalistic, with similar proportions across different background risk levels and risk communication formats.

TB15

Building A Level 4 A401

Data-Driven Aviation Operations

Contributed Session

Contributed

Chair: Adler Edsel, Purdue University, West Lafayette, IN, 47906, United States

1 - Air Cargo Tail Assignment and Shipment Routing Under Flight Time Uncertainty**Felipe Delgado, Pontificia Universidad Católica de Chile, Santiago, Chile, Shurui Zhu, Alessandro Bombelli**

Efficient and reliable air cargo transportation is critical for time-sensitive and high-value goods. However, disruptions caused by uncertain flight times present significant challenges, particularly when cargo requires tight connections. This paper presents a two-stage stochastic programming model that jointly optimizes aircraft tail assignment, flight schedule adjustments, and shipment routing under flight time uncertainty. The model accounts for heterogeneous cargo characteristics, including varying delay sensitivities and routing constraints, and incorporates operational penalties for delayed deliveries and flight rescheduling. To manage computational complexity, we introduce a tailored itinerary generation method and a scenario aggregation algorithm that preserves probabilistic characteristics while reducing solution time. The model is tested in an instance inspired by a major integrated logistics provider, including a night-time network of 86 flights among 35 Chinese airports. Computational experiments demonstrate that the proposed approach outperforms benchmark deterministic strategies based on expected and optimistic flight durations. The stochastic model achieves a 7.3% increase in average profit while enhancing load factors and reducing schedule disruptions. The results underscore the importance of integrating tactical and operational planning stages and accounting for uncertainty in air cargo logistics.

2 - Optimizing Baggage and Cargo Transport for the Aircraft Turnaround Process**Marcus Koh, Singapore University of Technology and Design, Singapore, Singapore, Rakesh Nagi, Nuno Ribeiro**

The aircraft turnaround process comprises a series of operations carried out between an aircraft's arrival at the gate and its subsequent departure. Given the involvement of multiple service vehicles performing diverse tasks, the Vehicle Routing Problem (VRP) provides a natural framework for modelling the process.

This research aims to develop a comprehensive model of turnaround operations to optimise resource planning across the airport. To generate a model that capture critical elements within the turnaround operations, a decomposition of the model into smaller VRP models for each separate GSE is considered. The first of the vehicle models in

focus is the baggage tractor vehicle that has several unique constraints that forces the consideration of multi-service, multi-trip, split delivery, heterogeneity, and intra-service precedence.

3 - Enhancing Airport Passenger Experience Through Large Language Models: Insights from Skytrax Data and Real-TIME Review Analytics

Burak Cankaya, Embry-Riddle Aeronautical University, Daytona Beach, FL, United States, Inci Polat Sesliokuyucu, Maria Petrescu, Unal Hayta, Scott Ambrose

Airports are working hard to provide better experiences for passengers, but it can be difficult to understand and respond to all the different needs and complaints travelers have. This study explores how Large Language Models (LLMs) and Knowledge Graphs, can help airports improve the way they serve passengers. By using data from Skytrax ratings and real-time Google reviews, we show how LLMs can quickly analyze large amounts of feedback to find common issues—such as confusing signs, long wait times, or unhelpful staff. We also developed a demo tool that creates visual maps of these insights, helping airports see where improvements are needed. We found that LLMs can make airport operations more efficient and help passengers feel more supported. We also provide a step-by-step guide for airports that want to start using AI in this way, along with ideas for future research.

4 - A Study of Scheduling and Fleet Routing for Urban Air Mobility Service

Adler Edsel, Purdue University, West Lafayette, IN, United States, William Crossley, Daniel DeLaurentis

Urban air mobility (UAM) is an emerging transportation system involving electric vertical-takeoff-and-landing (eVTOL) aircraft that operate with compatible infrastructure such as airports and heliports, collectively referred to as aerodromes. These aircraft carry passengers in and around metropolitan areas, with the intent of reducing travel times for these passengers compared to other modes of transport.

An on-demand UAM service would likely manifest as an aerial extension of current ground ride-hailing services (e.g., Uber and Lyft), whereas a scheduled UAM service would be modeled similarly to commercial airlines, involving a fixed timetable of flights and less dynamic pricing. Most existing UAM literature focuses on on-demand service; however, timetable-bound UAM operations may be a feasible alternative for a profit-seeking UAM operator if the schedule meets sufficient passenger demand. Implementing UAM operations bound to a fixed schedule may reduce the number of unprofitable deadhead flights required to satisfy demand for UAM services relative to an on-demand service model.

This study adapts airline scheduling and aircraft fleet routing methods to scheduled UAM operations and then uses these adapted methods to investigate a hypothetical UAM operator providing scheduled UAM services in a small three-aerodrome network serving the Chicago metropolitan area. A schedule design problem will determine a timetable of UAM flights informed by potential daily demand for UAM services, followed by a tail number assignment problem to assign aircraft throughout the network to meet this demand.

TB16

Building A Level 4 A402

Amazon Last Mile Planning

Invited Session

The Practice Section of Informs

Chair: Liron Yedidsion, Amazon, Redmond, WA, United States

1 - Fleet Allocation and Balancing of Branded Vehicles, Electric Vehicles and Rentals

Ram Thiruveedhi, Amazon, Austin, TX, United States, Abhilasha Katariya, Xiaodong Lan, Hugh Medal, Animesh Biyani

We present an optimization model for allocation and balancing of vehicle fleets to delivery stations. The Fleet Allocation and Balancing (FAB) model integrates heterogeneous fleet management across networks while accounting for operational constraints, costs, and sustainability requirements. Our approach unifies traditionally separate aspects of fleet management into a single optimization model. The FAB model optimizes multi-period planning horizons with weekly granularity, incorporating spatial distribution across stations and service partners, electric vehicle integration with charging infrastructure constraints, dynamic fleet composition through rental management and redeployments, and service partner requirements. The mathematical formulation captures the relationships between fleet availability, operational demands, and infrastructure limitations while maintaining computational tractability. We demonstrate the model's effectiveness using computational results showing cost reductions and efficiency improvements.

2 - Multi-Period Newsvendor for Capacity Planning

Liron Yedidsion, Amazon, Redmond, WA, United States, Chinmoy Mohapatra

Efficiently solving multi-period newsvendor problems is challenging due to the curse of dimensionality. We present a novel approach that combines the advantages of dynamic programming and reinforcement learning. By leveraging the discrete nature of the problem and using a neural network to approximate the value function, our method can handle high-dimensional state spaces without requiring a full simulation engine.

TB17

Building A Level 4 A403

AI and Optimization for Smarter Healthcare Systems

Panel Session

Practice Track

Co-Chair: Priya Sarathy, Wheel Data Strategies, 1171 Cumberland Road NE, Atlanta, 30306, United States

1 - Moderator Panelist

Priya Sarathy, Wheel Data Strategies, Atlanta, GA, United States

Healthcare systems around the world are under relentless pressure — grappling with rising costs, workforce shortages, care disparities, and increasingly complex patient needs. At the same time, advances in artificial intelligence (AI) and operations research (OR) offer the promise of smarter, faster, and more equitable solutions. But translating these tools from theoretical models to real-world impact remains a complex, unfinished journey.

This panel brings together cross-disciplinary leaders from healthcare, AI, and optimization to explore how data-driven decision-making can transform clinical operations, resource allocation, diagnostics, and care delivery. We will examine the hard questions: Why have intelligent scheduling systems failed to gain adoption despite clinician burnout? Can AI triage models be both efficient and ethical during crisis response? How do we optimize care pathways when data remains fragmented and biased? Panelists will share grounded perspectives from hospitals, research labs, startups, and policy circles. Topics will span predictive analytics for patient risk, prescriptive models for workforce planning, and the governance frameworks needed to build trust in AI at scale.

Rather than showcasing only success stories, this session will focus on real-world tension points—where theory meets resistance, and where opportunities for transformation lie just beyond the next iteration. Attendees will walk away with sharper insight into what it takes to operationalize intelligence in healthcare settings, along with key questions they should be asking as both practitioners and innovators.

2 - Panelist**Louis Luangkesorn, Highmark Health, Pittsburgh, PA, United States****3 - Panelist****Abrar Ahmed Syed, Gainwell Technologies, Irving, TX, United States****4 - Panelist****Deepti Bahel, KForce Inc, San Jose, CA, United States****TB18**

Building A Level 4 A404

Strategic Defense and Adversarial Decision-Making under Uncertainty

Invited Session

Military and Security

Chair: Isabella Sanders, Arizona State University, Newburgh, NY, United States

Co-Chair: Quanyan Zhu, New York University, New York, NY, United States

1 - The Location Set Covering Disruption Problem**Brian Lunday, Air Force Institute of Technology, WPAFB, OH, United States**, Andrew Sheldon, William Caballero

This research proposes a new, relevant competitive location problem and examines variants that relax complete and perfect game theoretic assumptions. The *location set covering disruption problem* (LSCDP) represents a Stackelberg game wherein the leader disrupts access to at most q out of n possible facility locations, after which the follower solves a perturbed location set covering problem (LSCP). Acknowledging the perturbed LSCP may not be feasible, the proposed LSCDP model represents a leader seeking to lexicographically maximize the value-weighted demands that cannot be covered and the follower's minimal cost of facilities required to cover remaining demands. Whereas the LSCDP adopts a complete and perfect game theoretic framework, two problem variants respectively account for incomplete and imperfect information: $q^c \leq q$ of the leader's disruptions are unknown to the follower, and a follower may be aware of the possibility of such covert disruptions but not their number. This research identifies appropriate solution methodologies for each of these variants and empirically demonstrates selected methods to derive relevant insights.

2 - Search for An Immobile Hider on a Binary Tree with Unreliable Locational Information**Thomas Lidbetter, Rutgers Business School, Newark, NJ, United States**, Steve Alpern

Adversarial search of a network for an immobile Hider (or target) was introduced and solved for rooted trees by Shmuel Gal in 1979. In this zero-sum game, a Hider picks a point to hide on the tree and a Searcher picks a unit speed trajectory starting at the root. The payoff (to the Hider) is the search time. In Gal's model (and subsequent investigations), the Searcher receives no additional information after the Hider chooses his location. In reality, the Searcher will often receive such locational information. For homeland security, mobile sensors on vehicles have been used to locate radioactive material stashed in an urban environment. In a military setting, mobile sensors can detect chemical signatures from land mines. In predator-prey search, the predator often has specially attuned senses that may help it locate the prey. How can such noisy locational information be used by the Searcher? We model such information as signals which indicate which of two branches of a binary tree should be searched first, where all signals have a known accuracy $p < 1$. Our solution calculates which branch (at every branch node) is favored, meaning it should always be searched first when the signal is in that direction. When the signal is in the other direction, we calculate the probability the signal should be followed. Compared with the optimal Hider strategy in the classic search game of Gal, the Hider's optimal distribution for this model is more skewed toward leaf nodes that are further from the root.

3 - Scheduling Cybersecurity Mitigations to Delay Attacker Projects**Ashley Peper, University of Wisconsin - Madison, Madison, WI, United States**, Laura Albert, Jim Luedtke

We study the problem of deploying cybersecurity mitigations to delay the completion times of a set of attacker projects. Existing literature studies how to select a portfolio of mitigations subject to a budget constraint to optimally delay such attack projects, but lacks consideration of the time it takes to implement the mitigations. We assume attackers are working on their projects at the same time the defender is working to complete mitigations to delay those attacks. Our proposed model captures the importance of scheduling mitigations, since to be useful they must be completed before an attacker exploits a vulnerability. This paper introduces an integer programming model of this problem that combines a defender's mitigation scheduling problem with the adversaries' project completion problem. We consider alternative solution methods and examine the benefit of the proposed model.

4 - Smart Tracking: How Hybrid UAV Scheduling and Kriging Enhance Fleet Surveillance**Esther Jose, University at Buffalo, Buffalo, NY, United States**, Moises Sudit, Rajan Batta

Tracking a fleet of targets is not the same as tracking multiple individual targets. This paper demonstrates how using surface resources to track a few vessels in a fleet can give us highly accurate probability heatmaps for the rest of the vessels in the fleet. This paper does so using a novel two-step framework that integrates a stochastic mixed-integer programming (MIP) model with spatio-temporal indicator Kriging (STIK) to optimize UAV assignments and dynamically update probabilistic heatmaps reflecting the locations of the rest of the fleet. First, the model schedules heterogeneous UAVs of two types: Synthetic Aperture Radar (SAR) to find the vicinity of a vessel, and Image Intelligence (IMINT) for precise identification, mimicking a tip-and-cue process. This is done to provide information to heatmaps that track the entire fleet.

The second stage employs STIK to assimilate these real-time sensor observations from the UAVs into dynamic heatmaps, capturing spatial and temporal autocorrelation among targets. Computational experiments, including detailed single-scenario analyses and larger randomized trials, demonstrate the robustness of our modeling framework, and show that piecewise-linear approximations optimally balance computational efficiency and solution accuracy. Although illustrated in maritime fleet tracking scenarios, our methodological contributions offer broad applicability to general fleet-based or group-target tracking problems prevalent in defense, search-and-rescue, and resource management operations.

TB19

Building A Level 4 A405

Decision Making with Limited Information: Optimization, Recommendations, Queues, and Matching

Invited Session

Applied Probability Society

Chair: Sushil Mahavir Varma, N/A

1 - Markov Decision Processing Networks**Neil Walton, Durham University, Durham, United Kingdom**

We introduce Markov Decision Processing Networks (MDPNs) as a multiclass queueing network model where service is a controlled, finite-state Markov process. The model exhibits a decision-dependent service process where actions taken influence future service availability. Viewed as a two-sided queueing model, this captures settings such as assemble-to-order systems, ride-hailing platforms, cross-skilled call centers, and quantum switches.

We first characterize the capacity region of MDPNs. Unlike classical switched networks, the MDPN capacity region depends on the long-run mix of service states induced by the control of the underlying service process. We show, via a counterexample, that MaxWeight is not throughput-optimal in this class, demonstrating the distinction

between MDPNs and classical queueing models.

To bridge this gap, we design a weighted average reward policy, a multiobjective MDP that leverages a two-timescale separation at the fluid scale. We prove throughput-optimality of the resulting policy. The techniques yield a clear capacity region description and apply to a broad family of two-sided matching systems.

2 - Impact of Rankings and Personalized Recommendations in Marketplaces

Akshit Kumar, Columbia Business School, New York, NY, United States, Omar Besbes, Yash Kanoria

Individuals frequently face a large set of alternatives while lacking a clear understanding of their own preferences. To facilitate choice, platforms deploy two ubiquitous information tools—public rankings and personalized recommendations—yet their comparative value across marketplace environments remains poorly understood. We develop a stylized model that spans two canonical settings: (i) uncapacitated supply, where any number of agents may select the same item, and (ii) capacitated supply, where each item can be matched to at most one agent. An agent's utility combines a common component that reflects population-level item quality with an idiosyncratic component that captures personal tastes; public rankings disclose only the common component, whereas personalized recommendations reveal both.

When supply is uncapacitated, either tool improves welfare relative to no disclosure. Public rankings are near-optimal when tastes are homogeneous, but personalized recommendations become increasingly valuable as heterogeneity grows because they exploit taste diversity. Under capacity constraints, revealing just the common component yields limited benefit: once an item's restricted common value is claimed, additional agents cannot share it. Personalized recommendations, by directing each agent to an item they idiosyncratically value highly, alleviate congestion and generate substantial welfare gains. The results demonstrate that the effectiveness of information provision is jointly governed by preference heterogeneity and supply constraints.

3 - Heavy-Traffic Analysis of Markov-Modulated Queues

Daniela Hurtado Lange, Northwestern University, Evanston, IL, United States, Izzy Grosof

In general, analyzing the delay distribution in queueing systems is intractable. However, key performance measures require knowledge of the distribution. For example, the service level requires computing the tail probabilities. To reduce this gap, a common practice in queueing theory is to analyze the system in an asymptotic regime, such as heavy traffic (where one loads the system close to its maximum capacity). Heavy traffic analysis has gained special attention in the last decade, when 'direct methods' have been developed. Here, one analyzes the steady state of the scaled queue lengths instead of a reflected Brownian motion. The Transform Method is one of them, and it focuses on studying the drift of an exponential transform of the scaled queue lengths (the test function) to show convergence in Moment Generating Function. A key step in this method is carefully designing a test function that yields a meaningful result. This method has been successfully used in systems where the inter-arrival and service times are i.i.d., and generalized to an input-queued switch with Markov-modulated arrivals. We propose using the Poisson equation to design the test function in this work. Specifically, we show that using the Poisson equation immediately yields the correct test function to study systems where the inter-arrival and service times depend on the queue lengths. Specifically, we study queueing systems with control in the arrival or the service process using the same essential steps. In this talk, I will present the methodology and examples of its use.

4 - Achieving Constant Regret for Dynamic Matching via State-Independent Policies

Mingwei Yang, Stanford University, Stanford, CA, United States, Süleyman Kerimov, Sophie Yu

We study a centralized discrete-time dynamic two-way matching model with finitely many agent types. Agents arrive stochastically over time and join their type-dedicated queues waiting to be matched. We focus on state-independent greedy policies that achieve constant regret at all times by making matching decisions based solely on agent availability across types, rather than requiring complete queue-length information. Such policies are particularly appealing for life-saving applications such as kidney exchange, as they require less information and provide more transparency compared to state-dependent policies.

First, for acyclic matching networks, we analyze a deterministic priority policy proposed by Kerimov et al. [2023] that follows a static priority order over matches. We derive the first explicit regret bound in terms of the general position gap (GPG) parameter ε —which measures the distance of the fluid relaxation from degeneracy.

Second, for general two-way matching networks, we design a randomized state-independent greedy policy that achieves constant regret with optimal scaling $O(\varepsilon^{-1})$, matching the existing lower bound established by Kerimov et al. [2024].

TB20

Building A Level 4 A406

AI and OR Applications in the Freight Railroad

Invited Session

Railway Applications

Chair: Nikola Bešinović, Technical University of Dresden, Dresden, Germany

Co-Chair: Daniel Haalboom, Technical University of Dresden, Dresden, Germany

1 - Optimizing Railroad Flat Yard Switching with Reward-Driven Integer Programming: a Sequential Decision Framework Using Branch-and-Bound and Remember Algorithm

Peiran Han, Beijing Jiaotong University, Beijing, China, People's Republic of, Lingyun Meng, Xiaojie Luan, Jianrui Miao, Zhengwen Liao, Nikola Bešinović

The Flat Yard Switching Optimization (FYSO) problem is a complex yet relatively underexplored challenge in railway freight operations. We introduce a Shunting Action Network (SAN) model, where each railcar is represented as a node and each shunting operation as a directed arc. To improve computational efficiency, we incorporate symmetry-breaking techniques and implement lazy constraints. Furthermore, we develop a sequential decision-making framework with a reward-driven structure to guide the optimization process. To enhance solution quality, we embed a memorization mechanism within a branching tree, leveraging state-based dominance and pruning strategies. Our approach supports various search strategies, including depth-first, breadth-first, best-first, and cyclic best-first search. The proposed method has been validated on a variety of case studies and demonstrates strong potential for extension to switching problems in broader operational contexts.

2 - Optimize Data-Driven and TIME-Varying Predictive Maintenance in Railway Networks

Yili Tang, University of Western Ontario, London, ON, Canada, Tangjian Wei

Accurate railway track gauge measurement is crucial for rail safety and efficiency. However, environmental exposure and complex track components in special areas such as turnouts and level crossings, introduce intrusions and noise interference, making precise measurement challenging. This paper presents a novel automated gauge measurement model based on the Spatial Density Clustering and Spatial Connection Constraint-Based Gauge Measurement (SDC-SCC-GM) Algorithm, utilizing LiDAR sensor data. The proposed SDC-SCC-GM algorithm integrates Spatial Density Clustering (SDC) and Spatial Connection Constraint (SCC) strategies to enhance measurement accuracy in complex railway environments. The SDC strategy employs Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to identify key points for gauge measurement, mitigating interference from noise points. Meanwhile, the SCC strategy expands the search space by placing potential key points into a candidate set and introducing spatial connection constraints to identify the most suitable key points for gauge measurement, effectively addressing track component interferences, particularly in turnouts and level crossings. To evaluate the effectiveness of SDC-SCC-GM algorithm, we conducted real-world field tests and compared our method with four baseline approaches. Experimental results demonstrate that SDC-SCC-GM outperforms all other methods, achieving the lowest error metrics across all rail conditions. The proposed method provides applicability and effectiveness of rail track gauge detection in complex railway environments.

3 - Railcar Shunting Optimization via An Adaptive Railcar Grouping Dynamic Programming Heuristic

Ruonan Zhao, Texas A&M University, College Station, TX, United States, Joseph Geunes, Xiaofeng Nie

Freight railyards commonly contain multiple tracks that require first-in-last-out (FILO) access for the classification and departure of railcars. Railcar Shunting Problems (RSPs) in such yards pose significant challenges because railcars with different destinations are mixed among different tracks, complicating retrieval and reassembly processes for outbound trains. This complexity is further intensified by the combinatorial explosion of possible railcar arrangements. To address these challenges, we propose an adaptive railcar grouping dynamic programming (ARG-DP) heuristic to solve the RSPs. Consecutive railcars sharing the same destination are merged into unified groups, ensuring that these groups move together without splitting. Additionally, ARG-DP adaptively groups railcars with common destinations at each shunting movement and employs a reduced distance-to-go function to construct the dynamic programming graph, significantly reducing the size of the network. Computational experiments demonstrate the effectiveness of ARG-DP in simulated and real-world railyards. On average, across 60 test cases of simulated yards, ARG-DP achieves solutions 355 times faster than an exact mixed-integer programming model, maintaining an average optimality gap of 6.65%. Moreover, in 10 test cases based on the Gaia yard in Portugal, ARG-DP obtains solutions 229 times faster on average, with an optimality gap of 6.90%.

4 - Strategic Planning of Classification Yard Operations Under Infrastructure and Service Constraints

Daniel Haalboom, Dresden University of Technology, Dresden, Germany

Classification yards are essential components of Single Wagonload (SWL) rail networks, enabling efficient transitions between inbound and outbound railcars. However, increasing competition from other transport modes and structural changes in the European industrial landscape necessitate a strategic redesign of these networks. The absence of historical operational data complicates long-term planning, while the high capital cost of yard infrastructure demands efficient use. Simultaneously, certain customer requirements necessitate rapid railcar interchange on specific routes. These competing demands require careful selection of which direct transitions to support, as not all can be accommodated. Railcars not assigned a direct transition must either traverse mixed tracks with subsequent rehumping, or bypass the hump entirely, requiring additional shunting operations. To address this trade-off, we propose a mixed-integer linear programming (MILP) model for the strategic scheduling of operations within classification yards. The model focuses on optimizing railcar transitions, prioritizing time-sensitive cars while minimizing the use of mixed tracks and re-humping operations. It explicitly incorporates infrastructure constraints and internal yard connections to ensure operational feasibility and avoid conflicts. We evaluate the proposed model using real-life-scale instances of classification yards. The results yield detailed, car-specific movement patterns and provide actionable insights to support strategic decision-making in SWL network design.

TB21

Building A Level 4 A407

Market Design and Network Games (I)

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Manxi Wu, University of California, Berkeley, Berkeley, CA, United States

1 - Information Design for Strategic Driver Reposition on Ride-Hailing Platforms

Manxi Wu, University of California, Berkeley, Berkeley, CA, United States, Ozan Candogan

In this paper, we study how ride-hailing platforms like Uber and Lyft can strategically share information about the demand distribution with drivers to influence repositioning behavior. On these platforms, drivers have the flexibility to move strategically to different locations for better payoffs, and the spatial distribution of service requests and drivers is crucial for platform efficiency and revenue. We model drivers' strategic repositioning decisions as a game. Using a Bayesian persuasion framework, we show that in many practically relevant cases, the optimal mechanism takes a simple threshold form: the demand realization is fully revealed when it is below or above certain thresholds, and not revealed in the intermediate region. We also develop a computational method to solve for optimal information mechanisms in general settings, and extend the framework to the joint design of information and pricing.

2 - Integrated Equilibrium Model for Electrified Logistics and Power Systems

Rui Yao, EPFL, Lausanne, Switzerland, Xuhang Liu, Anna Scaglione, Shlomo Bekhor, Kenan Zhang

This paper proposes an integrated equilibrium model to characterize the complex interactions between electrified logistics systems and electric power delivery systems. The model consists of two major players: an electrified logistics operator (ELO) and a power system operator (PSO). The ELO aims to maximize its profit by strategically scheduling and routing its electric delivery vehicles (e-trucks) for deliveries and charging, in response to the locational marginal price (LMP) set by the PSO. The routing, delivery, and charging behaviors of e-trucks are modeled by a perturbed utility Markov decision process (PU-MDP) while their collective operations are optimized to achieve the ELO's objective by designing rewards in the PU-MDP. On the other hand, PSO optimizes the energy price by considering both the spatiotemporal e-truck charging demand and the base electricity load. The equilibrium of the integrated system is formulated as a fixed point, proved to exist under mild assumptions, and solved for a case study on the Hawaii network via Anderson's fixed-point acceleration algorithm. Along with these numerical results, this paper provides both theoretical insights and practical guidelines to achieve sustainable and efficient operations in modern electrified logistics and power systems.

3 - Congestion Pricing Revisited: Balancing Efficiency and Equity in Urban Transportation

Qianni Wang, Northwestern University, Evanston, IL, United States, Xinlei Zhou, Jun Xie, Marco Nie

Congestion pricing is a popular tool for travel demand management. However, it is often criticized for being regressive, i.e., benefiting the wealthy at the expense of the poor. Here, we attempt to address this concern for fairness through mechanism design. Building on recent advances in inequality-aware market design, we analyze a stylized network model to characterize inequality-aware congestion pricing mechanisms. We show that an optimal mechanism can take one of three forms: (i) a deterministic posted-price scheme, (ii) a lottery-based rationing scheme, or (iii) a hybrid mechanism combining a higher deterministic price with a lower lottery price granting access with limited probability. Leveraging the structural properties of these optimal mechanisms, we develop a numerical algorithm to facilitate their design and examine their efficiency (e.g., the price of anarchy) and welfare effects (e.g., Gini coefficient). We also discuss the potential extension to general transportation networks.

4 - A Game Between Ridehailing Operator and Government Under the Incentive Bundle Program

Jason Lu, University of Michigan, Ann Arbor, MI, United States, Neda Masoud

Multimodal transportation addresses the first-last mile problem by integrating fixed-route transit services with flexible or on-demand travel modes to support complete user journeys. However, compared to direct car trips, multimodal journeys can be more time-consuming or costly, discouraging users from adopting the system. In this paper, we consider a setting where travelers rely on public transit as the primary mode of transportation, using ridesourcing or walking to bridge the first and last miles of their journeys. To support and encourage the adoption of multimodal transportation, we propose an Incentive Bundle Program (IBP), administered by the government, which allocates subsidies to participating users to help offset ridesourcing costs. We develop a game-theoretic approach to capture the interactions among three key stakeholders in the IBP: the government, the ridehailing operator, and users. The government determines the subsidy distribution, the ridesourcing operator sets pricing, and users choose their preferred modes of transportation based on their individual utilities. We propose a tri-level optimization framework to model these interactions and transform the problem into a single-level convex program. For large-scale instances, we design a solution methodology that achieves strong solution quality while maintaining computational efficiency. Using real-world data, we conduct numerical studies to analyze how government subsidy allocation and ridehailing pricing decisions impact user behavior in the multimodal transportation system.

TB22

Building A Level 4 A408

Frontiers in LLM and AI agent

Invited Session

Applied Probability Society

Chair: Tianyi Peng, Columbia University, 541 West 113th Street, New York, NY, 10025, United States

1 - Speculative Actions: a Lossless Framework for Faster AI Agents

Naimeng Ye, Columbia University, New York, NY, United States

Despite growing interest in AI agents across industry and academia, their execution in an environment is often slow, hampering training, evaluation, and deployment. For example, letting two state-of-the-art agents play a game of chess may take hours. A key bottleneck is that agent behavior unfolds sequentially: each action requires an API call, and these calls can be time-consuming. Inspired by speculative execution in microprocessors and speculative decoding in LLM inference, we propose Speculative Actions, a lossless framework that predicts likely actions using faster models, enabling multiple steps to be executed in parallel. We evaluate this framework across four agentic environments: gaming, e-commerce, web search, and operating systems. In all cases, speculative actions yield substantial acceleration, with potential speedups of up to 30%. Moreover, performance can be further improved through stronger guessing models, top-\$K\$ action prediction, multi-step speculation, and uncertainty-aware optimization, opening a promising path toward real-world, efficient deployment of AI agents.

2 - Optimizing LLM Inference: Fluid-Guided Online Scheduling with Memory Constraints

Gan Luo, Peking University, Beijing, China, People's Republic of, Ruicheng Ao, David Simchi-Levi, Xinshang Wang

Large Language Models (LLMs) is indispensable in today's applications, but their inference procedure—generating responses by breaking text into smaller pieces and processing them using a memory-heavy element named Key-Value (KV) cache—requires a lot of computational resources, especially when memory is limited. This paper treats LLM inference optimization as a multi-stage online scheduling problem, where prompts arrive sequentially and the incremental expansion of the KV cache during inference renders conventional scheduling algorithms ineffective. To address this challenge, we develop fluid dynamics approximation to establish a tractable benchmark, providing insights for devising effective scheduling algorithms. Building upon this foundation, we introduce the Waiting for Accumulated Inference Threshold (WAIT) algorithm as a warm-up. This method maintains multiple thresholds to determine the scheduling order of incoming prompts, optimizing resource utilization when output lengths are known at the time of arrival. In practical applications where output lengths are not known at the time of prompt arrival, we extend our method by introducing the Nested WAIT algorithm. This algorithm constructs a hierarchical framework comprising multiple segments, each defined by distinct thresholds, to effectively manage the random prompt arrivals with unknown output lengths. Theoretical analysis shows both algorithms have near-optimal performance compared with the fluid benchmark under heavy traffic limit, balancing throughput, latency, and Time to First Token (TTFT). Numerical experiments conducted with the Llama-7B model on an A100 GPU, utilizing both synthetic and real-world datasets, demonstrate that our approach achieves superior throughput and reduced latency compared to widely adopted baseline methods such as vLLM and Sarathi.

3 - Llm Inference Optimization via Operations Research

Jim Dai, Cornell University, Ithaca, NY, United States, Tianze Deng, Yueying LI, Tianyi Peng

As demand for Large Language Models (LLMs) and AI agents rapidly grows, optimizing systems for efficient LLM inference becomes critical. While significant efforts have targeted system-level engineering, little is explored through a mathematical modeling and queueing perspective. In this paper, we aim to develop the queueing fundamentals for LLM inference, bridging the gap between queueing and LLM system communities. In particular, we study the throughput aspect of LLM inference systems. We prove that a large class of 'work-conserving' scheduling algorithms can achieve maximum throughput for both individual requests and AI-agent workloads, highlighting 'work-conserving' as a key design principle in practice. Evaluations of real-world systems show that Orca and Sarathi-serve are throughput-optimal, reassuring practitioners, while FasterTransformer and vanilla vLLM are not maximally stable and should be used with caution. Moreover, under memory constraints, we identify a non-trivial necessary condition for stability, yielding practical insights into hyperparameter selection for Sarathi-serve. Our results highlight the substantial benefits queueing community can offer in analyzing and improving LLM inference systems and call for more interdisciplinary developments.

4 - What is Your AI Agent Buying? Evaluation, Implications, and Emerging Questions for Agentic E-Commerce

Amine Allouah, MyCustomAI, Santa Clara, CA, United States, Omar Besbes, Josué D Figueroa, Yash Kanoria, Akshit Kumar

Online marketplaces will be transformed by autonomous AI agents acting on behalf of consumers. Rather than humans browsing and clicking, vision-language-model (VLM) agents can parse webpages, evaluate products, and transact. This raises a fundamental question: what do AI agents buy—and why? We develop ACES, a sandbox environment that pairs a platform-agnostic VLM agent with a fully programmable mock marketplace to study this question. We first conduct basic rationality checks in the context of simple tasks, and then, by randomizing product positions, prices, ratings, reviews, sponsored tags and platform endorsements, we obtain causal estimates of how frontier VLMs actually shop. Models show strong but heterogeneous position effects: all favor the top row, yet different models prefer different columns, undermining the assumption of a universal "top" rank. They penalize sponsored tags and reward endorsements. Sensitivities to price, ratings, and reviews are directionally human-like but vary sharply in magnitude across models. Motivated by scenarios where sellers themselves use AI agents to optimize product listings, we show that a seller-side agent that makes minor tweaks to product descriptions—targeting AI buyer preferences—can deliver substantial market-share gains if AI-mediated shopping dominates. We also find that modal product choices can differ across models and, in some cases, demand may concentrate on a few select products, raising competition questions. Together, our results illuminate how AI agents may behave in e-commerce settings, and surface concrete seller strategy, platform design, and regulatory questions in an AI-mediated ecosystem.

TB23

Building A Level 4 A409

Theory and Applications of Simulation

Contributed Session

Contributed

Chair: Xin Fei, University of Edinburgh, Edinburgh, EH11 2NH, United Kingdom

1 - Rate-Optimal Strategy for Best Policy Identification in Reinforcement Learning

Mingjie Hu, Fudan University School of Management, Shanghai, China, People's Republic of, Edward Huang, Jie Xu, Jianqiang Hu, Chun-Hung Chen

Best policy identification (BPI), which seeks to identify the policy with the highest expected cumulative reward, has recently gained considerable attention in reinforcement learning (RL). This paper formulates a BPI problem for finite-horizon, time-inhomogeneous, tabular Markov decision processes (MDPs) within a fixed-budget ranking and selection (R&S) framework, with the goal of maximizing the probability of identifying the best policy. To tackle this problem, we first derive a rate-optimal sampling ratio using large deviation theory, incorporating the dynamics and rewards of MDPs into the rate function. We then quantify the benefits of incorporating MDP structural information and show that the optimal sampling ratio balances variance, optimality gap, and visitation probabilities. Motivated by abundant covariate information in practical decision-making, we propose a contextual BPI framework with linear reward functions and extend the optimal sampling ratio to this setting. In addition, we introduce an asymptotically rate-optimal strategy and an efficient approximate strategy based on a closed-form sampling ratio. Numerical experiments demonstrate that our proposed algorithm outperforms existing R&S and RL methods. Furthermore, we illustrate its practical applicability through a complex diffusion furnace scheduling problem. These results highlight the potential of R&S techniques for solving BPI problems.

2 - The Price of Streaming

Franco Berbeglia, Purdue University, West Lafayette, IN, United States, Timothy Derdenger, Sridhar Tayur

We study the price of streaming strategies on total movie revenues. Using panel data from 2014–2021 on US box office ticket sales, transactional video-on-demand (TVOD) rentals, digital sales, DVD/Blu-ray sales, and rentals, as well as streaming video-on-demand (SVOD) availability information, we estimate a dynamic structural model of demand across channels. We first analyze a direct-to-streaming strategy implemented by Warner Bros. (WB) and HBO Max: Project Popcorn. We determine that WB was not incorrect with its streaming strategy. We provide quantitative support and qualitative insights on the performance of hybrid strategies: simultaneous release in theaters and streaming, and delayed streaming within the theatrical window. Through "what-if" analyses our structural model can aid studio executives in crafting their streaming strategy by analyzing the pertinent trade-offs.

3 - Unifying CGE and Network Equilibrium Models of Commodity Demand by Industrial Sector

Patrick O'Reilly, Demand Scientific, Golden, CO, United States

Estimating commodity demand within specific economic sectors requires capturing both economy-wide market interdependencies and sector-specific network dynamics. This paper proposes a novel integration of Computable General Equilibrium (CGE) models, formulated as Mixed Complementarity Problems (MCPs), with Network Equilibrium (NE) models, expressed as Variational Inequalities (VIs), to forecast demand for a commodity in a targeted sector. By exploiting the mathematical equivalence of VIs and MCPs, I unify the CGE's market-clearing conditions with the NE's network flow constraints into a single MCP. Alternatively, the model can be formulated as a VI, preserving the NE's structure while appending CGE equations as constraints. A simplified example model demonstrates the approach, modeling commodity demand driven by sector-specific network interactions. The unified model captures substitution effects, price-quantity dynamics, and network feedbacks, offering richer insights than unlinked CGE and NE models that otherwise speak to the same context. Computational experiments, implemented in Julia/JuMP, demonstrate the model's scalability and convergence for monotone problems. Sensitivity analyses begin with exploring demand responses to price variation, offering implications for relative scarcity. This work extends prior CGE-NE integrations by focusing on commodity-specific demand, addressing a gap in sector-targeted applications. Presented results aim to motivate development into dynamic, multi-commodity settings.

4 - Efficient Bayesian Sampling Policies for Fixed-Budget Ranking and Selection

Xin Fei, University of Edinburgh, Edinburgh, United Kingdom, Juergen Branke

We study the fixed-budget ranking and selection problem, in which a decision maker must identify the best design from a finite set using a limited number of samples. The challenge lies in balancing exploration - allocating samples to designs not currently identified as best but characterised by significant uncertainty or the potential to outperform the current leader - and exploitation, focusing sampling efforts on designs presently estimated as best. This paper introduces a novel allocation policy designed to minimize either the probability of incorrect selection or the expected opportunity cost. Grounded in large deviation theory, the policy sequentially allocates samples to the design that potentially provides the largest value of sample information. We provide a theoretical analysis and conduct comparative empirical evaluations against selected state-of-the-art methods.

TB24

Building A Level 4 A410

Modeling with Attack Graphs for Securing Cyber-physical Systems

Tutorial session

Tutorials

Chair: Brian Denton, University of Michigan, Ann Arbor, MI, United States

1 - Modeling with Attack Graphs for Securing Cyber-physical Systems

Laura Albert, University of Wisconsin-Madison, Madison, WI, United States, Carmen Haseltine

Modern critical infrastructures consist of increasingly complex and interdependent cyber-physical systems (CPS). Securing cyber-physical infrastructure requires understanding how components interact with one another across physical, cyber, and human dimensions, as well as how threat modeling affects system components and operations. This tutorial reviews modeling techniques for CPS security, focusing on network models, threat modeling techniques, and prescriptive decision-making. Attack graphs offer a structured and systematic approach to modeling threats to a system and are crucial in cybersecurity risk management. We review three applications of operations research modeling techniques that leverage attack graphs for CPS security: allocating a security budget for cybersecurity planning, developing vulnerability metrics for securing cyber-physical energy systems, and performing risk assessments for administering election systems. Through these applications, we demonstrate the potential for using attack graphs for prescriptive decision-making and proactive planning to secure CPS.

TB25

Building A Level 4 A411

Integration of Learning and Routing

Invited Session

TSL: Freight Transportation

Chair: Dan Liu, Kean University, Union, NJ, United States

1 - CombFormer: Learn to Separate Strengthened Comb Inequalities for Cvrp

Merve Doganbas, KAIST, Daejeon, Korea, Republic of, Hayong Shin, Changhyun Kwon

The Capacitated Vehicle Routing Problem (CVRP) is a fundamental NP-hard problem in logistics and transportation. A key exact method for solving CVRPs is the branch-and-cut algorithm, which depends on the separation of strong cutting planes to tighten linear programming (LP) relaxations. Among these, Strengthened Comb Inequalities (SCIs) form a class of facet-defining cuts that incorporate vehicle capacity constraints into comb inequalities for the Traveling Salesman Problem. We propose CombFormer, a hybrid heuristic for separating SCIs using a learning-based approach. It formulates the SCI handle construction as a sequential decision-making problem over the LP solution. A transformer-based graph neural network encodes the fractional solution, and an attention-based decoder autoregressively selects handle nodes. Once the handle is formed, a greedy teeth-construction subroutine completes the SCI. The policy is trained with REINFORCE to maximize the violation of the generated cut. Training is conducted on instances with 50–100 nodes, and the model is fine-tuned on larger ones (150–300 nodes). In experiments, we compare our method to SCI separation heuristic of CVRPSEP, a cut separation algorithm package that is widely used for solving VRPs exactly. CombFormer performs comparably to CVRPSEP on small instances and significantly outperforms it on larger problems (up to 500 nodes), identifying SCIs with higher violations. These results suggest that structurally complex inequalities like SCIs can be effectively separated using learning-based techniques trained to recognize polyhedral structures. CombFormer complements classical heuristics by demonstrating the feasibility of learning-guided cut generation for complex inequalities.

2 - Test-TIME Search in Neural Graph Coarsening Procedures for the Capacitated Vehicle Routing Problems

Yoonju Sim, KAIST, Daejeon, Korea, Republic of, Hyeonah Kim, Changhyun Kwon

Recent advances in deep learning have motivated various efforts to enhance cutting-plane methods by integrating learned models. Kim et al. (2024) introduced a neural separation (NeuralSEP) framework that utilizes graph neural networks to generate rounded capacity inequalities (RCIs) for capacitated vehicle routing problems (CVRP). In this paper, we refine and extend NeuralSEP in two key directions: (1) by advancing its implementation and (2) by extending it to produce framed capacity inequalities (FCIs) without further training. Through our analysis of NeuralSEP, we identify a critical limitation: small input variations lead to poor cut diversity, reducing the effectiveness of the separation approach. To address this, we introduce stochastic edge selection by adding small random perturbations to edge contraction probabilities, enabling the generation of more diverse cuts. Also, we migrate the implementation from DGL to PyTorch Geometric, reducing computational overhead during graph coarsening operations. Furthermore, we leverage insights from the coarsening steps in NeuralSEP to generate FCIs at test time by examining the mapping history to identify promising candidate partitions. Experiments on randomly generated instances of CVRP show that our enhanced NeuralSEP scales more effectively than the original NeuralSEP and CVRPSEP, reducing computational time while achieving optimality gap reductions of 1-2% on average for instances with more than 300 nodes.

3 - A Contextual Online Prize-Collecting Vehicle Routing Problem

Amirmohammad Paksaz, University of Iowa, Iowa City, IA, United States

We consider a setting in which the decision maker coordinates a vehicle fleet to serve customers but has limited or no initial data on each customer's value. In each period, the decision maker (i) observes related contextual information, (ii) leverages this information in the form of feature variables to infer information about the stochastic customer prizes, (iii) determines a set of routes that visit a subset of the customers, (iv) collects the observed prizes, and (v) updates the prize-learning model. Thus, the decision-maker faces the well-known "exploration-exploitation" trade-off to efficiently allocate limited delivery resources across a spatially distributed customer base with an unknown context-dependent demand distribution that must be learned incrementally. We represent our contextual online prize-collecting vehicle routing problem as a combinatorial multiarmed bandit in which each arm corresponds to a customer. At each round, the decision maker selects a subset of arms (customers) that can be feasibly served by a given fleet of vehicles. We study a semibandit setting where we observe the prizes associated with each arm in the selected subset but do not observe the prizes associated with unselected arms. We seek a learning algorithm that minimizes the time-average pseudoregret. We present preliminary results obtained by modifying existing multiarmed bandit algorithms to our setting.

4 - Metaroutenet: a Meta-Reinforcement Learning Framework for Adaptive Electric Vehicle Routing with TIME Windows

Mertcan Daysalilar, University of Miami, Miami, FL, United States, Fuat Uyguroglu, Ramiro Valdes Jara, Gabriel Nicolosi, Adam Meyers

The Electric Vehicle Routing Problem with Time Windows (EVRPTW) introduces routing complexities due to energy constraints, charging infrastructure, and delivery time windows. Traditional optimization methods often lack flexibility in adapting to new problem variations, requiring manual redesign or tuning, while current reinforcement learning approaches typically struggle to generalize and must be retrained for each new scenario—making both approaches less suitable for dynamic logistics environment. This work proposes MetaRouteNet, a meta-reinforcement learning framework that combines Model-Agnostic Meta-Learning (MAML) and Proximal Policy Optimization (PPO) for fast adaptation across EV routing tasks. Each routing instance is modeled as a separate task, allowing MetaRouteNet to learn generalizable knowledge through meta-training. A graph-attention encoder and joint vehicle-node attention mechanism model spatiotemporal interactions between customers, charging stations, and vehicle states. To improve policy robustness, we employ curriculum learning, gradually exposing the model to increasingly complex customer scenarios (e.g., 5-, 10-, and 15-node problem instances). We are currently evaluating MetaRouteNet on the Schneider EVRPTW dataset, which includes instances across multiple complexity levels. Preliminary results indicate that MetaRouteNet may offer advantages over baseline PPO and static policy models in metrics such as adaptation time, route feasibility, and total distance. We anticipate that its meta-learning architecture will generalize effectively to unseen instances with varying customer demands and time constraints. These early findings suggest that MetaRouteNet is a promising solution for EV fleet management in dynamic, real-world environments. By bridging reinforcement learning and combinatorial optimization, this research contributes to adaptive, energy-efficient routing systems.

5 - Optimizing Urban Grocery Delivery: a Deep RL Framework with Linformer Attention and a Joint Decoder for Heterogeneous Fleet Routing

Dan Liu, Kean University, Union, NJ, United States

The Heterogeneous Capacitated Vehicle Routing Problem (HCVRP) represents a pivotal challenge in urban grocery delivery. This is due to the need to optimize routes for a diverse fleet with varying capacities, speeds, and cost structures. This heterogeneity adds layers of complexity that go beyond the traditional CVRP. Moreover, as the size of the problem increases (i.e., more nodes and vehicles), traditional optimization methods—whether exact algorithms or heuristics—often face exponential increases in computational demands. To address these challenges, this study introduces a novel neural network architecture that incorporates advanced attention mechanisms tailored specifically for HCVRP with backhaul considerations. Our approach features two key innovations. First, we integrate the Linformer, which employs a low-rank approximation attention mechanism to significantly reduce the computational demands of conventional Transformer models, thereby improving efficiency on large datasets. Second, we develop a Multi-Relational Joint Selection Decoder that dynamically updates node representations in real time, enabling more precise routing decisions based on current vehicle positions, previously traveled paths, and inbound backhaul requirements. Extensive experiments reveal that our deep reinforcement learning (DRL) framework consistently outperforms traditional heuristics and existing DRL methods in terms of solution quality, convergence speed, and computational efficiency across diverse problem scales and objectives. This research underscores the transformative potential of integrating state-of-the-art machine learning techniques with classical operations research methodologies, paving the way for more resilient, efficient, and sustainable transportation and logistics solutions in complex urban environments.

TB26

Building A Level 4 A412

Simulation as a Digital Twin

Invited Session

Simulation

Chair: Canan Gunes Corlu, Boston University, Boston, MA, United States

1 - Digital Twins: Applications and Future Directions

Canan Gunes Corlu, Boston University, Boston, MA, United States, Bahar Biller, Stephan Biller

Digital twins are an essential concept for organizations pursuing digital transformation. In this talk, we explore the benefits of digital twin technology for supply chains, including improved supply chain resilience. We also examine the challenges associated with implementing digital twins and highlight opportunities for future research.

2 - Real-Time Digital Twin Approach for Personalized Medical Decision Making

Yangyuechen Zhao, Boston University, Boston, MA, United States, John Maleyeff, Fangzhou Xie

Digital twin approaches are a relatively new concept in healthcare. To support shared decision making, we utilize the digital twin model to create a virtual representation of a patient who needs to choose among a candidate list of treatment options. The digital twin exists within a real-time decision support system (DSS) that integrates electronic health records (EHR) and patient-reported outcomes (PRO). The DSS includes a decision aid that visualizes treatment benefits and side effects with their corresponding probabilities. Using a Bayesian framework, the digital twin continuously updates likelihood predictions for each individual patient based on their personal characteristics, medical histories, and health conditions. Unlike traditional digital twins used in supply chain or engineering domains that tend to be simulations. Here, a healthcare-focused digital twin consists of a prediction model that creates a prior prediction distribution. Bayesian updating is done using data from the EHR and PRO to create the posterior prediction distribution. In these cases, sample sizes are often small. An example is presented concerning carotid endarterectomy surgery with the possibility of stroke as a side effect. Medicare data is used to create a prediction model that serves as the prior prediction model, where the prediction is assumed to follow a beta distribution. Two clinical scenarios illustrate how the digital twin logic either creates the posterior prediction model or alerts the medical staff when PRO data differ significantly from predictions.

3 - Bridging the Physical and Virtual: Real-Time Simulation Modeling for Next-Generation Digital Twins

Ghaith Rabadi, University of Central Florida, Orlando, FL, United States, Bulent Soykan

Digital Twins (DTs) are rapidly transforming how complex systems are designed, monitored, and managed. A critical enabler for realizing the full potential of DTs, particularly in dynamic operational environments, is the integration of real-time simulation modeling. This presentation explores novel methodologies and frameworks for embedding high-fidelity, real-time simulation capabilities within DT architectures. We discuss how operations research techniques, including dynamic data assimilation, adaptive model calibration, and distributed simulation, can address the challenges of synchronizing virtual models with their physical counterparts in real-time. The talk will also address computational considerations and future research directions in this burgeoning field at the intersection of simulation, AI, and management science.

4 - Simulation Digital Twin Validation with Multi-Epoch Key Performance Indicators

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

We propose a validation framework to verify whether the distributions of the simulated and system-observed key performance indicators (KPIs) are identically distributed across multiple epochs. The KPIs are transient and non-identically distributed across the epochs, which makes the validation challenging as we do not have repeated measurements of the system KPIs. To tackle this problem, we propose a hypothesis testing framework that consists of a set of marginal tests and one joint inter-correlation

test. Each marginal test corresponds to each KPI and examines if the marginal distribution of the Digital Twin (DT) and system KPI within each epoch align in all observed epochs. For each epoch, the system KPI is converted via probability integral transform using the empirical marginal cumulative distribution function of the simulated KPIs. The converted marginal KPIs from all epochs are then tested against a standard uniform distribution. If the marginal tests fail to reject, the joint test determines whether the inter-correlation of the DT and system KPIs, measured by their respective copula functions, match. Assuming that the copula structure remains unchanged across the epochs, we pool the multi-epoch data and apply nonparametric copula estimates based on the transformation used in the marginal tests. A copula hypothesis testing is then performed to compare the DT and system KPIs. We establish asymptotic analyses to ensure that the tests deliver a valid Type-I error. To further improve the finite-sample performance, we adopt a stepdown multiple hypothesis testing framework where a set of critical values are calculated using bootstrap.

TB27

Building B Level 2 B201

Interplay of Optimization and Statistics in ML

Invited Session

Computing Society

Chair: Yifan Hu, Rutgers, Piscataway, NJ, United States

1 - Transformers Provably Learn Chain-of-Thought Reasoning with Length Generalization

Yu Huang, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States, Zixin Wen, Aarti Singh, Yuejie Chi, Yuxin Chen

The ability to reason lies at the core of artificial intelligence, and tough problems usually call for deeper, more extended reasoning to tackle. A crucial question about AI reasoning is whether models can extrapolate from learned reasoning patterns to solve harder tasks that require longer chain-of-thoughts (CoT). In this work, we present a theoretical analysis of transformers trained via gradient descent on synthetic data for various state tracking tasks, revealing how length-generalizable reasoning can emerge. Specifically, we prove that 1) for tasks with simple algebraic structure, e.g., composition in a cyclic group, transformers learn a solution pattern that generalizes to polynomial-length CoT; 2) for more complex tasks like composition in a symmetric group, a recursive self-training curriculum enables transformers to bootstrap longer reasoning and achieve polynomial generalization. Intuitively, our results demonstrate that transformers can learn sequential reasoning skills that scale with problem complexity. Moreover, we provide the first provable optimization-based guarantee for constant-depth transformers on the state tracking problems beyond circuit complexity class TC^0 : we prove that they can learn problems in NC^1 , which goes beyond prior results that consider problems within TC^0 , unless the famous conjecture $\mathsf{TC}^0 \neq \mathsf{NC}^1$ is false.

2 - Causal Invariance Learning via Efficient Optimization of a Nonconvex Objective

Zhenyu Wang, Rutgers University, New Brunswick, NJ, United States, Yifan Hu, Peter Bulhmann, Zijian Guo

Identifying causal relationships among variables from observational data is important yet challenging. Data from multiple environments provide opportunities to uncover causality by exploiting the invariance principle that the causal outcome model should remain the same across heterogeneous environments. In this paper, we consider the problem of identifying the direct parents of the outcome from a set of covariates and learning their effects, i.e., to learn the causal outcome model. We first propose a novel Negative Weighted Distributionally Robust Optimization (NegDRO) to learn an invariant prediction model. Under the additive interventions regime, we establish three major contributions. 1) On the statistics side, we provide nearly necessary and sufficient identification conditions under which the invariant prediction model coincides with the causal outcome model. 2) On the optimization side, despite nonconvexity of NegDRO, under these identification conditions, we establish the benign nonconvex landscape of NegDRO, i.e., any stationary points are globally optimal, paving the path for computation. 3) On the computational side, we further develop a gradient descent maximization method that provably converges to the causal outcome model, and establish the convergence rate with respect to the sample size and iteration times. Our method avoids exhaustive combinatorial search over all possible (exponential many) subsets of covariates or integer programming used in the invariance learning literature, ensuring scalability when the number of covariates is large. To our knowledge, this is the first causal invariance method that guarantees global optimality of a nonconvex objective. Numerical results further validate the efficiency of the proposed method.

3 - Robustness and Regret Tolerance

Yilin Xue, National University of Singapore, Singapore, Singapore, Napat Rujeerapaiboon

This paper shifts focus from the typical approach of maximizing expected reward to minimizing expected regret, aiming to find a solution whose expected reward is close to the oracle. While these two approaches are equivalent when the uncertainty distribution is given, they diverge when accounting for distributional ambiguity, which is characterized by the Wasserstein distance, for enhanced robustness. This can be achieved by requiring the decision maker to specify their tolerance for expected regret under the nominal distribution and minimize the excess regret beyond this threshold across any other distribution. By exploring the entire range of this threshold parameter, our regret-based model encompasses both empirical optimization and worst-case regret minimization, the latter of which is NP-hard. When the reward function is represented by a linear program (possibly involving recourse decisions, a random technology matrix, and a random right-hand side), we propose two tractable safe approximations that preserve feasibility and empirical consistency of the original problem. The first approximation is formulated as a robust linear program with convex uncertainty sets, while the second tightens and improves upon the first using positive semidefinite constraints. Finally, we demonstrate the advantages of our tolerance-oriented regret-based model over a similar target-oriented reward-based model and empirical optimization in the context of supply chain procurement and portfolio selection.

4 - Preconditioning for Overparameterized and Ill-Conditioned Composite Optimization

Abdel Ghani Labassi, Johns Hopkins University, Baltimore, MD, United States, Mateo Diaz, Liwei Jiang

Composite optimization problems involve minimizing the composition of a smooth map with a convex function. Such objectives arise in numerous data science and signal processing applications, including phase retrieval and blind deconvolution. The subgradient method achieves local linear convergence when the composite loss is well-conditioned. However, if the smooth map is, in a certain sense, ill-conditioned or overparameterized, the subgradient method exhibits much slower sublinear convergence even when the convex function is well-conditioned. To overcome this limitation, we introduce a novel Levenberg-Marquardt subgradient method that converges linearly under mild regularity assumptions at a rate determined solely by the conditioning of the convex function. Further, we demonstrate that these regularity assumptions hold for several problems of practical interest, including square-variable formulations, low-rank matrix, and tensor recovery problems. Numerical experiments illustrate the benefits of our method.

5 - Doubly Fair Policy Learning

Zeyu Bian, Florida State University, Tallahassee, FL, United States

Fairness and trustworthiness have recently received significant attention in machine learning, particularly in supervised learning, driven by ethical concerns surrounding algorithmic applications. However, fairness in policy learning--where the goal is to make fair decisions (actions) while simultaneously maximizing expected rewards (value) has been seldomly explored. Most existing works in fairness for policy learning focus primarily on action fairness. Yet, in many applications, ensuring fairness in both actions and value is critical. This paper tackles the challenge of double fairness in policy learning by addressing three objectives simultaneously: action fairness, value fairness, and maximizing the value function. We propose a novel framework that integrates fairness constraints directly into the multi-objective optimization problem for policy learning. Our framework is flexible and accommodates various fairness notions. To the best of our knowledge, this is the first work to address double fairness in policy learning. We establish theoretical guarantees for the proposed method and demonstrate its empirical superiority over competing approaches.

TB28

Building B Level 2 B202

Federated Learning and Optimization I

Invited Session

Computing Society

Chair: Kibaek Kim, Argonne National Laboratory, Lemont, IL, United States

Co-Chair: Farzad Yousefian, Rutgers University, Piscataway, NJ, United States

1 - Differentially Private Federated Averaging with James-Stein Estimator

Minseok Ryu, Arizona State University, Tempe, AZ, United States, Geunyeong Byeon, Xinran Zhao

Differentially private federated learning (DPFL) enables collaborative model training without centralizing sensitive data, ensuring privacy via formal differential privacy mechanisms. A key challenge in DPFL is improving model utility under strong privacy constraints. To address this, we propose DP-FedStein, a variant of the differentially private federated averaging (DP-FedAvg), which incorporates the James-Stein estimator (JSE) at the server to more accurately estimate true pseudo-gradients from their noisy, differentially private versions. We provide a convergence analysis showing that, by leveraging the superior performance of JSE in high-dimensional settings, DP-FedStein achieves a better utility-privacy trade-off than DP-FedAvg. Empirical results on public and private image classification datasets confirm its effectiveness, demonstrating substantial accuracy gains under high noise conditions.

2 - Scaling Federated Learning to HPC: A Cross-Silo Framework for Real-World Large-Scale Model Training

Yijiang Li, Argonne National Laboratory, Lemont, IL, United States, Kibaek Kim

High-performance computing (HPC) systems offer unprecedented computational resources for federated learning (FL), yet deploying FL across heterogeneous HPC environments presents unique infrastructure and implementation challenges. This work presents a comprehensive framework for conducting real-world cross-silo federated learning experiments across diverse supercomputing facilities including Polaris, Aurora, Perlmutter, and Frontier. Our framework addresses these deployment challenges by extending the APPFL package: Globus Compute Endpoints for distributed task orchestration across heterogeneous architectures, Proxystore and Globus Transfer for efficient model synchronization and data movement, and DeepSpeed integration for multi-node parallel training. This enables training of large-scale applications including large language models, computer vision tasks, and time series forecasting. We demonstrate the framework's effectiveness through comprehensive benchmarks of various FL algorithms, providing practical insights into scalability, convergence behavior, and deployment considerations. Our results advance large-scale federated learning by enabling realistic experiments that mirror real-world collaborative machine learning scenarios across the HPC ecosystem while maintaining data sovereignty.

3 - Adding and Removing Clients on the Fly: A Federated Learning Approach

Krishnan Raghavan, Argonne National Laboratory, Lemont, IL, United States, Kibaek Kim, Todd Munson

Federated learning (FL) is one of the most common distributed learning paradigms that allow the flexibility of training large scale models with a multitude of clients with heterogeneous data. However, typical FL does not fully support dynamic addition and removal of clients. In this talk, we seek to illuminate the difficulties introduced due to on the fly addition/removal of clients and provides potential solutions to this problem. In particular, we will discuss a probabilistic viewpoint of FL that is naturally poised to address this challenge. We will demonstrate theoretically and practically the viability of this approach and develop experimental results supporting our claims.

4 - A Randomized Zeroth-Order Hierarchical Framework for Heterogeneous Federated Learning

Yuyang Qiu, Rutgers University, Piscataway, NJ, United States, Kibaek Kim, Farzad Yousefian

Heterogeneity in federated learning (FL) is a critical and challenging aspect that significantly impacts model performance and convergence. In this talk, we propose a novel framework by formulating heterogeneous FL as a hierarchical optimization problem. This new framework captures both local and global training process through a bilevel formulation and is capable of the following: (i) addressing client heterogeneity through a personalized learning framework; (ii) capturing pre-training process on server's side; (iii) updating global model through nonstandard aggregation; (iv) allowing for nonidentical local steps; and (v) capturing clients' local constraints. We design and analyze an implicit zeroth-order FL method (ZO-HFL), provided with nonasymptotic convergence guarantees for both the server-agent and the individual client-agents, and asymptotic guarantees for both the server-agent and client-agents in an almost sure sense. Notably, our method does not rely on standard assumptions in heterogeneous FL, such as the bounded gradient dissimilarity condition. We implement our method on image classification tasks and compare with other methods under different heterogeneous settings.

5 - Adaptive Federated Learning Across Heterogeneous High-Performance Computing Systems

Kibaek Kim, Argonne National Laboratory, Lemont, IL, United States, Yijiang Li

High-performance computing (HPC) systems commonly rely on job schedulers to manage resource allocation by queuing and executing jobs based on requested computing time and capacity. When conducting federated learning (FL) experiments across multiple, heterogeneous HPC systems, additional challenges arise due to the variability in system architectures, job scheduling policies, and resource availability. In this work, we identify and discuss these challenges, emphasizing their impact on performance and scalability in cross-system FL training. To address these issues, we propose a novel FL algorithm that actively monitors local system status and adaptively adjusts resource usage at each client. This adaptive strategy enables more efficient and robust training under the constraints imposed by diverse HPC environments. We demonstrate preliminary results from applying our method on U.S. Department of Energy (DOE) supercomputers, showing promising improvements in resource utilization and training efficiency.

TB29

Building B Level 2 B203

Advanced Machine Learning for the Future Power Grid

Invited Session

ENRE: Energy

Chair: Mucun Sun, Idaho National Laboratory, Idaho Falls, 83401, United States

Co-Chair: Yinan Wang, Rensselaer Polytechnic Institute, 110 8th St, Troy, 12180-3522, United States

1 - Imitation Learning for Distribution and Microgrid Intelligence: from Self-Healing Grid to Performance Guaranteed Control

Yichen Zhang, University of Texas at Arlington, Arlington, TX, United States

Modern deep learning has been strengthening the capability of automatic decision-making and intelligent control under complex scenarios and has become an essential piece of the blueprint for distribution intelligence and smart grids. Because of the strong representation power, a deep neural network (DNN) can make better and faster decisions to reroute power flow through network reconfiguration, especially under multiple outage scenarios, to facilitate self-healing distribution systems. DNN-based dynamic controllers can encode disturbance estimation, state and input constraint satisfaction, and expert knowledge into an end-to-end policy. All these advances rely on the construction of intelligent agents, and the reinforcement learning framework is naturally suitable for this task. However, there are currently two major limitations. First, traditional reinforcement learning algorithms are exploration dominant and exhibit low training efficiency, especially under random disturbances. Secondly, for dynamic control tasks, current training paradigms cannot ensure closed-loop stability. In this presentation, we will demonstrate a new training paradigm, a synergy between imitation learning, mathematical programming, and control theory, to improve the training efficiency and ensure system stability under random disturbances. The presentation investigates two problems: automatic network reconfiguration with reactive power dispatch for service restoration of distribution systems and frequency control with stability and safety guarantees in islanded microgrids. The successful results indicate that the imitation learning framework provides an opportunity to leverage well-established mixed-integer programming solvers and control theories for learning-based automations.

2 - Physics-Informed Online Learning for Dynamic State Estimation in Multimachine Power Systems

Guanyu Tian, Texas A&M University, Galveston, TX, United States

With the increasing penetration of distributed generators in the smart grids, having knowledge of rapid real time electromechanical dynamic states has become crucial to system stability control. Conventional Supervisory Control and Data Acquisition (SCADA)-based dynamic state estimation (DSE) techniques are limited by the slow sampling rates, while the emerging phasor measurement units (PMUs) technology enables rapid real-time measurements at network nodes. Using generator bus terminal voltages, we propose a hybrid-learning DSE (HL-DSE) algorithm to estimate the synchronous machine rotor angle and speed in real time. The HL-DSE takes the power system model into account and trains neuro-estimators with real-time data in an online manner. Compared with traditional DSE methods, the HL-DSE overcomes limitations by using a data-driven approach in conjunction with the physical power system model. The time efficiency, accuracy, convergence, and robustness of the proposed algorithm are tested under noise and fault conditions in both small- and large-scale test systems. Simulation results show that the proposed HL-DSE is much more computationally efficient than widely used Kalman filter (KF)-based methods while maintaining comparable accuracy and robustness. In particular, HL-DSE is over 100 times faster than square-root unscented KF (SR-UKF) and 80 times faster than extended KF (EKF). The advantages and challenges of the HL-DSE are also discussed.

3 - Repower: Enabling Autonomous Research in Power Systems with Large Language Models

Yongxin Zhang, University of Tennessee, Knoxville, TX, United States

Large Language Models (LLMs) have shown strong potential across various scientific fields, yet their application in power systems remains limited. In this talk, I will present Revive Power Systems (RePower), an autonomous research platform that employs a reflection-evolution strategy to independently address complex power system challenges. RePower autonomously controls devices, gathers data, designs methods, and evolves algorithms, achieving consistent improvements across tasks such as parameter prediction, power optimization, and state estimation. This work demonstrates a new direction for autonomous innovation in power systems research.

4 - Efficient and Effective Contrastive Self-Supervised Learning Framework for Power Time Series Data

Yifeng Gao, University of Texas Rio Grande Valley, Edinburg, TX, United States

Contrastive Self-Supervised Learning (CSSL) based training framework is widely used to train deep neural networks. By leveraging the ability to learn meaningful representations through view augmentation instead of task-specific labels, the CSSL framework trains a pre-training encoder model that can capture general data properties and is suitable for diverse downstream tasks. In the field of smart energy, CSSL framework is especially beneficial given the existence of diverse sensor signals and the challenge of accessing detailed human annotations. However, due to the rich information entangled within sensor signals, existing CSSL frameworks often train models that store an enormous amount of information in the embedding space, which significantly reduces model efficiency. In this talk, we will discuss the rationale and recent advancement on effective CSSL training framework for power time series. We will also discuss and introduce methods to enhance the inference efficiency.

5 - Physics-Informed Machine Learning Towards Cyber-Resilient Distribution Grid Situational Awareness

Ying Zhang, Oklahoma State University, Stillwater, OK, United States

Modern power systems are confronted with degraded and even stranded operational performances subject to growing complexity, uncertainty, and volatility from large-scale integration of distributed energy resources (DERs). Real-time distribution grid situational awareness must be obtained from data of heterogeneous sensors to support intelligent control and decision-making. This talk will present a new technique for AI-aided grid operation by developing physics-informed trustworthy learning for fast yet accurate state estimation in electric distribution systems with massive DERs—combining power engineering knowledge with a data-aided regression technique. Unlike black-box data-intensive machine learning that has limited creditability in real-world grids, the proposed method achieves significant performance synergy in terms of accuracy, efficiency, and robustness against outliers or cyber-attacked measurements. As system-wide voltage and power flow visibility is of fundamental importance, the proposed framework has the potential as a building block to accelerate and advance various system-wide optimization tasks, such as voltage control, thereby digitally twinning the physical grid.

TB30

Building B Level 2 B204

Advancements in Generation and Transmission Expansion Planning

Invited Session

ENRE: Electricity

Chair: Todd Levin, Argonne National Laboratory, Lemont, IL, 60439, United States

1 - Deployment Pathways for Long Duration Energy Storage

Todd Levin, Argonne National Laboratory, Lemont, IL, United States, W. Neal Mann, Jonghwan Kwon, Zhi Zhou, Bo Cheng, Audun Botterud

We apply a least-cost generation expansion model of the continental United States to assess how optimal investments in long-duration energy storage (LDES) technologies are impacted by changes in system generation portfolios and technology costs, assessing 180 capacity expansion scenarios in total. The expansion model considers 8,760 hours of chronological operations for the entire target year, 2040. We find that low-cost LDES technologies can reduce generation investments and system costs. Specifically, once costs for 24-hour and 100-hour storage reach \$38/kWh and \$14/kWh, respectively, substantial deployments are observed. The distribution of storage investments across durations is strongly influenced by the system generation portfolio. We also demonstrate that a high-fidelity temporal representation is required to capture the value of LDES in generation expansion. Finally, we find that LDES deployments in the generation expansion are positively correlated with the combined wind and solar capacity share and negatively correlated with peaking and baseload shares.

2 - Regional Household-Level Impacts of Long-Term Transmission Strategies

Charalampos Avraam, National Renewable Energy Laboratory, Golden, CO, United States, Maxwell Brown, Yasser El Masri, Sarah Awara

Transmission infrastructure can allow consumers in remote regions to access the electricity generation of regions with more abundant resources, reducing the system-level cost and enhancing regional and system-level affordability and reliability. However, the costs and benefits of transmission expansion and resource availability can differ based on the type of household. The existing literature explores transmission and resource availability conditions for diverse households by assessing the operational implications of electricity disruptions. However, there is little emphasis on the role of long-term transmission expansion strategies and resource availability when accounting for the response across diverse households. Moreover, collection and processing of household-level data across the U.S. is extremely resource-intensive. Besides, large-scale national studies integrating the behavioral response of consumers by household type at the sub-state level can be computationally intensive. In this study, we leverage and process building-level electricity consumption data from ResStock across all U.S. households to generate inverse demand curves for five different household types at the sub-state level. Then, we integrate the inverse demand curves into the Regional Energy Deployment System (ReEDS), a large-scale capacity expansion model of the electric power sector of the contiguous U.S., to explore the impact of transmission expansion and regional resource availability to diverse U.S. consumers. By identifying critically affected regional households, this work aims to inform policymakers and regional stakeholders engaged in long-term transmission infrastructure design towards ensuring affordable, reliable, and abundant energy for all through the mid-century.

3 - Transmission System Approximations for Nodal Transmission and Generation Co-Expansion

Alyssa Kody, National Renewable Energy Laboratory, Golden, CO, United States, Bernard Knueven, Elizabeth Glista, Jean-Paul Watson

To support increasing and shifting customer loads and to uphold federal and state energy policy, the co-expansion problem jointly selects new generation and new transmission over a multi-decade planning horizon. System planners rely on various simplifications of the co-expansion problem for computational tractability. Examples of these simplifications include: (1) collapsing a network into zones or using a copper-plate model, (2) ignoring transmission physics using a “pipe-and-bubble” power flow approximation, and (3) the use of representative days or load flow cases as a proxy for a decades-long time horizon. Although these types of simplifications are widely used, little work has been performed to evaluate how closely these simplifications reflect the more detailed models that are used in operation. The focus of this paper is to evaluate example (2) and explore the quality of investment decisions when using simplified transmission models, like “pipe-and-bubble”, for nodal representations of realistically sized networks.

4 - Large-Scale Robust and Adaptive Planning for Generation and Transmission Expansion: California Case Study

Gabe Mantegna, Princeton University, Princeton, NJ, United States

Electric utility planners and regulators must frequently make decisions about long-lived infrastructure under deep uncertainty surrounding cost, demand, and technology availability. However, most planning practitioners use scenario analysis alone for considering uncertainty, a planning process which can frequently miss out on infrastructure decisions that offer much greater flexibility and robustness to uncertainty. Methods exist for identifying robust and adaptive decisions in an optimization framework, but they have not yet been applied to the specific context of realistic-scale generation and transmission expansion planning. In this study we develop a method for applying two-stage adaptive robust optimization to realistic-scale generation and transmission expansion planning for the state of California, and demonstrate its advantages relative to today's deterministic planning practices.

5 - Long-Term Capacity Planning Under Operational Uncertainty

KE XIN ZUO, Cornell University, ITHACA, NY, United States, Jacob Mays

This paper investigates how operational policies for managing uncertainty in power systems affect long-term capacity investment decisions in wholesale electricity markets. Traditional capacity expansion models (CEMs) typically assume perfect foresight and static representations of system dynamics, failing to capture the variability, forecast uncertainty, and intertemporal constraints present in real-world operations. Such simplifications result in the systematic mis-valuation of flexible resources (e.g. battery storage.) While theoretically, socially-optimal prices can be determined from extreme-scale stochastic programs with extensive scenario trees, such approaches remain computationally infeasible for practical implementation. In reality, system operators face a sequential decision-making problem where market-clearing prices from the operational stage coordinate capacity investments through a zero-profit condition in equilibrium. Consequently, market prices become dependent on the specific formulation and parameterization of the algorithms implemented to manage operations. Using a fixed-point iteration algorithm, we simulate operations under various policy assumptions—including Stochastic Lookahead Unit Commitment (SLAC) and Deterministic Lookahead Unit Commitment (DLAC) variants—and iteratively update capacity investments until convergence to the zero-profit equilibrium. Our results demonstrate that different operational policies yield fundamentally different price formations and capacity equilibria, despite identical underlying market structures. This algorithmic dependency of prices introduces critical inefficiencies in current market designs, distorting investment incentives and potentially leading to a long-run resource mix that diverges from the theoretical ideal. We conclude with recommendations for aligning operational uncertainty management with efficient long-term investment signals to improve resource adequacy.

TB31

Building B Level 2 B205

Renewable Integration and Storage Systems

Contributed Session

Contributed

Chair: luce brotcorne, INRIA, Lille, France

1 - Shifting Renewable Energy Generation Under Climate Change

Renee Obringer, Pennsylvania State University, University Park, PA, United States

As the climate crisis intensifies, switching to renewable energy remains a critical piece of the solution to ensure rapid decarbonization. However, renewable energy generation is highly reliant on the ambient environmental conditions, making it difficult to estimate the long-term generation—a task that is likely to get more difficult under climate change. Accounting for the impact of climate change is particularly difficult, as there remains uncertainty related to the magnitude of climate change within the mid- and long-term in addition to the relatively unknown impacts of climate change on generation of renewable energy technologies. In this work, we aim to fill this gap by leveraging machine learning to investigate the impact of climate change on state-level renewable energy generation across the US. Using data from the Energy Information Administration (EIA), we project the solar, wind, and hydropower generation across multiple US states under two key climate change scenarios. Our goal is to answer two key questions: (1) How will climate change impact renewable energy generation; and (2) Do these impacts differ across states? To answer these questions, we leveraged several machine learning techniques, as well as an ensemble of models, to first model the observed relationship between renewable energy generation and the surrounding weather and climate. Then, we used those same models to project the changes to the system, given the most recent IPCC climate change scenarios.

2 - Optimized Co-Production of Cement and Methanol for Cost-Efficient Decarbonization

Yuezhang He, Tsinghua University, Beijing, China, People's Republic of, Jesse Jenkins

Cement and methanol production together contribute approximately 7.3% of global energy-related CO₂ emissions, underscoring their significance in achieving deep decarbonization. Given the substantial technical and economic barriers to separately decarbonizing these hard-to-abate sectors, we propose a renewable-powered co-production system that utilizes molecule exchange—captured CO₂ and by-product O₂—to reduce emissions more cost-effectively. We develop an hourly-resolved, process-based optimization model that incorporates operational flexibility and applies it to plant-level deployment strategies and CO₂ source-sink matching across China. Our results show that co-production can lower CO₂ abatement costs to \$41–53 per tonne by 2035, compared to around \$75 for standalone cement CCUS and over \$120 for renewable-based methanol. Introducing flexible methanol synthesis further reduces costs by 2.9%–16.6% through improved load balancing and storage integration. The co-production system also facilitates CCUS adoption in cement plants located in regions with abundant renewable energy, indicating a strategic opportunity to reshape national CO₂ transport and storage infrastructure. This study offers a system-level framework for jointly decarbonizing the cement and methanol industries, leveraging technological synergies and spatial heterogeneity to enhance cost-effectiveness and deployment feasibility.

3 - Resilient Renewable-Powered Multi-Functional Charging Station for Freight Transport at Maritime Ports

Jeremiah Gbadegoye, University of Tennessee, Knoxville, TN, United States, Yang Chen, Olufemi Omitaomu, Xueping Li

As freight transportation undergoes a paradigm shift toward decarbonization, ports emerge as strategic nodes for integrating clean energy infrastructure into intermodal logistics systems. This study proposes a simulation-based framework for optimizing and evaluating the operations of a Multifunctional Charging Station (MFCS) at a maritime port, capable of serving electric and hydrogen-powered freight vehicles including trucks, trains, and ships. The MFCS integrates renewable generation (solar photovoltaics and wind turbines) alongside hydrogen production via electrolysis and grid interaction to enable flexible, low-emission freight operations. We explore infrastructure capacity planning under uncertainty, focusing on capacity sizing for battery storage, hydrogen storage, and electrolyzers, using both historical freight data and forward looking scenarios based on Freight Analysis Framework (FAF) projections. The model incorporates real-world developments in battery-electric ships and inland waterway freight to ground technology adoption trends. Importantly, we examine system resilience by evaluating performance under grid outages, renewable variability, and evolving transport demands, comparing trade-offs between hydrogen and battery storage expansion. By simulating adoption trajectories of clean freight fleets through 2050, we offer a dynamic view of energy and logistics interdependencies at ports. The results inform investment strategies that balance operational costs, carbon impact, and infrastructure robustness. This research provides a reproducible approach to support planners and policymakers in scaling MFCS deployment at port facilities, emphasizing the critical role of renewable-integrated, resilient charging infrastructure in enabling the future of sustainable freight.

4 - Cloud Sharing for Energy Efficiency

luce brotcorne, INRIA, Lille, France, Nathalia Wolf

Cloud computing has revolutionized IT infrastructure by providing scalable and cost-efficient solutions. However, the rising demand for cloud services has intensified concerns around resource inefficiency and energy consumption. In this talk we propose a dynamic pricing and resource-sharing model designed to optimize virtual machine (VM) utilization. By reallocating idle VMs from long-term users to short-term consumers, this approach significantly improves resource efficiency while lowering energy costs.

Computational experiments demonstrate the effectiveness of this model in enhancing VM utilization and reducing energy consumption. Our results highlight the importance of incentivizing lender participation for maximizing efficiency and maintaining profitability in cloud services.

TB32

Building B Level 2 B206

Power Grid Resilience Against Extreme Events

Invited Session

ENRE: Energy-Climate

Chair: Matthew Brun, Operations Research Center, MIT, Cambridge, MA, United States

Co-Chair: Xu Sun, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Quantifying Cascading Power Outages During Climate Extremes Considering Renewable Energy Integration**Luo Xu, Princeton University, Princeton, NJ, United States**

Climate extremes, such as hurricanes, combined with large-scale integration of environment-sensitive renewables, could exacerbate the risk of widespread power outages. We introduce a coupled climate-energy model for cascading power outages, which incorporates grid operations and comprehensively captures the impacts of climate extremes on renewable generation, and transmission and distribution networks. The model is validated with the 2022 Puerto Rico catastrophic blackout during Hurricane Fiona – a unique system-wide blackout event with complete records of weather-induced outages. The model reveals a resilience pattern that was not captured by the previous models: early failure of certain critical components enhances overall system resilience. Sensitivity analysis on various scenarios of behind-the-meter solar integration demonstrates that lower integration levels (below 45%, including the current level) exhibit minimal impact on system resilience in this event. However, surpassing this critical level without pairing it with energy storage can exacerbate the probability of catastrophic blackouts.

2 - Optimizing Battery and Line Undergrounding Investments for Transmission Systems Under Wildfire Risk Scenarios: a Benders Decomposition Approach**Ryan Piansky, Georgia Institute of Technology, Atlanta, GA, United States, Daniel Molzahn, Rahul Gupta**

With electric power infrastructure posing an increasing risk of igniting wildfires under continuing climate change, utilities are frequently de-energizing power lines to mitigate wildfire ignition risk, which can cause load shedding. Recent research advocates for installing battery energy storage systems as well as undergrounding risky overhead lines to reduce the load shedding during such de-energizations. Since wildfire ignition risk can exhibit substantial geographic and temporal variations, it is important to plan battery installation and line undergrounding investments while considering multiple possible scenarios. This paper presents a scenario-based framework for optimizing battery installation and line undergrounding investments while considering many scenarios, each consisting of a day-long time series of uncertain parameters for the load demand, renewable generation, and wildfire ignition risks. This problem is difficult to solve due to a large number of scenarios and binary variables associated with the battery placements as well as the lines to be undergrounded. To address the computational challenges, we decompose the problem in a two-stage scheme via a Benders decomposition approach. The first stage is a master problem formulated as a mixed-integer linear programming (MILP) model that makes decisions on the locations and sizes of batteries as well as the lines to be undergrounded. The second stage consists of a linear programming model that assesses these battery and line undergrounding decisions as modeled by a DC optimal power flow formulation. We demonstrate the effectiveness of the proposed scheme on a large transmission network with actual data on wildfire ignition risks, load, and renewable generation.

3 - Designing Wildfires for Adversarial Power Grid Outages**Matthew Brun, Operations Research Center, MIT, Cambridge, MA, United States, Xu Sun, Jean-Paul Watson**

Electric power infrastructure faces increasing risks of damage and disruption due to wildfires. The planning and operation of electric grids in wildfire-prone regions must consider the potential impact of unpredictable fires. However, traditional wildfire models do not easily answer how and when wildfires may disable some set of power grid elements. To address these questions, we propose an optimization model that designs an adversarial wildfire to target specific infrastructure while respecting physical fire spread dynamics. We identify a convex set for wildfire spread and formulate the adversarial model as a mixed-integer nonconvex program. We apply a novel convexification and dual reformulation that yield an infinite-dimensional mixed-integer conic program, which is solved via variable and constraint generation. We apply our approach to produce worst-case counterfactual fires that disable power grid elements near the recent Park and Eaton fires in California.

4 - An (Extreme) Probabilistic Perspective into Battery Trilemma: to Charge, to Charge, OR to Wait?**Tomas Tapia, Johns Hopkins University, Baltimore, MD, United States, Yury Dvorkin**

This presentation develops a risk-aware framework for positioning energy-limited battery storage under wholesale electricity market uncertainty. We first consider multi-interval battery dispatch using a finite-horizon stochastic dynamic formulation with a probabilistic end-of-horizon state-of-charge (SoC) constraint, ensuring with a given confidence level that the SoC remains within a target value. Building on this formulation, we derive explicit buy/sell thresholds using the k-search approach, with provable worst-case competitive ratios, and compute corresponding action probabilities across the finite horizon. To accelerate computations, we introduce two pruning techniques, probability redistribution and path pruning, and leverage them to quantify the likelihood of satisfying the end-of-horizon SoC target. Finally, we benchmark our results against a tractable robust optimization counterpart that incorporates budgeted uncertainty sets and propagation methods to estimate worst-case stopping times. Computational experiments using historical market prices show that the proposed framework significantly improves arbitrage revenues while maintaining system reliability and safety.

5 - Decision-Dependent Uncertainty-Aware Distribution System Planning Under Wildfire Risk**Alexandre Moreira, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, Felipe Pianco, Bruno Fanzeres, Ruiwei Jiang, Chaoyue Zhao, Miguel Heleno**

The interaction between power systems and wildfires can be dangerous and costly. Distribution grids can be liable for the outbreak of wildfires during extreme weather. In wildfire-prone areas, investment planning should consider the impact of operational actions on wildfire-related uncertainties affecting line failure likelihood. In this case, endogenous-based uncertainty modeling should comprise the backbone of the investment planning model viz-a-viz the inability of standard exogenous-based uncertainty modeling. Therefore, we propose a decision-dependent uncertainty (DDU) aware methodology to optimize investment portfolios for distribution systems, considering that high power-flow levels in high-threat areas can ignite wildfires and increase line failure probability. The methodology identifies the best combination of upgrades (new lines, hardening existing lines, and placing switching devices). Methodologically, we propose a two-stage distributionally robust planning optimization problem with DDU that considers the distribution system's multiperiod operation. The first stage determines optimal switching actions and line investments, and the second stage evaluates the worst-case expected operational cost under a DDU framework designed to account for the endogenous impact of power-flow levels and hardening investment decisions in the line failure probabilities. An iterative method is tailored to handle the problem and numerical experiments demonstrate a more prepared grid to deal with wildfire risk.

TB33

Building B Level 2 B207

Analytical Models in Healthcare

Invited Session

Health Applications Society

Chair: Hossein Piri, Haskayen School of Business-University of Calgary, Calgary, AB, Canada

1 - MI Compass: Pareto Frontier-Based Navigation of MI Model Trade-Offs**Vasileios Digalakis, Boston University, Boston, MA, United States, Ramayya Krishnan, Agni Orfanoudaki**

We propose a structured framework for ML model selection that systematically balances internal performance metrics (e.g., accuracy, interpretability, fairness) with external evaluation criteria (e.g., adoption and regulatory compliance). The framework formalizes the roles of modelers, users, and regulators, embedding their objectives within an optimization problem that maximizes the decision-maker utility under real-world constraints.

2 - Cost-Effective Surveillance Strategies for Early Detection of Cancer Recurrence: Optimal Integration of Imaging and ctDNA Testing

Reza Skandari, Imperial College Business School, London, United Kingdom, Narges Mohammadi, Alexander Pearson

Patients with cancer remain at risk of recurrence, either locoregionally or via distant metastasis, following completion of primary treatment. Timely identification of recurrence is essential for initiating treatment early and improving survival outcomes. Current surveillance strategies primarily rely on imaging, which are costly and associated with high false-positive rates leading to unnecessary biopsies and impose significant psychological and physical burdens on patients.

Circulating tumor DNA (ctDNA) testing offers a promising blood-based alternative that is more cost-effective and demonstrates higher accuracy. However, it lacks the ability to localize the site of recurrence. To address this limitation, we propose a cost-effective surveillance strategy that integrates both imaging and ctDNA testing.

We develop a constrained partially observable Markov decision process (CPOMDP) model to optimize surveillance costs under different target detection delays. Our framework introduces a Bayesian aggregation approach to personalize surveillance strategies across six patient subgroups, stratified by cancer stage and smoking habit.

Our findings indicate that the integrated strategy achieves detection delays comparable to current guidelines while significantly reducing overall costs—resulting in an estimated \$300 million in annual savings—and lowering false-positive rates from 56.5% to 6.5%. This work presents a scalable, data-driven framework for optimizing post-treatment surveillance in head and neck cancer, with potential applicability to other cancer types and future non-invasive diagnostic innovations.

3 - Convexity in Transient Queues with Appointment-Driven Arrivals and Implications on Dynamic Scheduling

Christos Zacharias, Miami Herbert, Miami, FL, United States, Alex Kuiper

We propose a unified platform to assess stochastic convexity in appointment scheduling. This paper relates to both intra-day and inter-day scheduling. In continuous intra-day scheduling, stochastic convexity is crucial in ensuring that minimization routines reach a global minimum - the optimal static schedule for a single day. We introduce to the scheduling literature the concept of strong stochastic convexity, which facilitates the assessment of convexity of performance measures of interest across a diverse range of stochastic settings. Moreover, our novel dynamic programming model encompasses inter-day scheduling, allowing us to jointly consider the assignment of both times and days for patients' appointments. This integration of intra-day and inter-day scheduling enables us to account for indirect delays and waitlist considerations when making appointment decisions.

4 - Pooling Uncommon Incidences Among Hospital Monitoring Locations for Decision-Making or Among Research Studies for Meta-Analyses

Franklin Dexter, University of Iowa, Iowa City, IA, United States, Rakesh Sondekoppam, Emine Bayman

Hospital engineering quality monitoring involves rare events (e.g., postoperative infections). Also, anesthesia procedural complications, such as neurological injury from peripheral nerve blocks, are rare. Consider combining postoperative infection data among adjacent operating rooms (e.g., for consideration of a common daily intervention such as ultraviolet-C disinfection). The economically important endpoint is the overall infection incidence (i.e., marginal estimate [e.g., 2%]). There is complete data, with room, numerator of infections, and denominator of cases. Several earlier Monte-Carlo simulation studies examined the accuracy of confidence interval coverage and bias when combining such incidences, guiding suitable analyses. However, these assume that the logit or probit of the incidence or rate follows normal distributions, which is not true in practice. Our simulation studies show marked impacts on results. Software produces nice-looking control charts, but inaccurate. Consider, similarly, meta-analyses for incidences of anesthesia complications. Investigators textually look at calculated confidence intervals and make statements about differences among methods and changes over time. For example, "the incidence of neurological dysfunction was approximately 0.9% (95% CI 0.8% to 1.1%) at <2 weeks' follow-up to approximately 0.03% (95% CI 0.01% to 0.05%) at 1 year." There is inherently complete data for meta-regression of incidences because each study provides a numerator, a denominator, and a study characteristic (e.g., follow-up period). Using generalized linear models with probit or logit link and single fixed effects variables (e.g., year), or generalized estimating equations, we again find highly inaccurate confidence interval coverage. The statistics packages converge, giving answers, but often highly unreliable.

TB34

Building B Level 2 B208

Advances in Machine Learning

Invited Session

OPT: Machine Learning

Chair: Brian Liu, MIT, Cambridge, MA, 02139, United States

1 - Foundations and Applications in Explainable AI

Brian Liu, MIT, Cambridge, MA, United States

Tree ensembles, such as random forests and gradient boosting, remain among the most effective algorithms for structured, tabular data, and consistently outperform deep learning methods in these settings. However, their opacity make them challenging to use in high-stakes domains such as healthcare and public policy, where model explainability is crucial. In this talk, I will present a general optimization-based framework to extract explainable models from black-box tree ensembles. Our framework is flexible and allows practitioners to balance predictive accuracy with criteria such as sparsity, interaction depth, transparency, and stability. The extracted models are accurate and trustworthy, and can be examined to reveal useful relationships in the data.

2 - Benign Overfitting Beyond Prediction

Dennis Shen, USC, Los Angeles, CA, United States

Deep learning research has uncovered the benign overfitting phenomenon for overparameterized statistical models, which has drawn significant interest in recent years. Given its simplicity and practicality, the ordinary least squares (OLS) interpolator has become essential to gain foundational insights into this phenomenon. While the properties of OLS are well established in underparameterized settings, its behavior in overparameterized regimes is less explored (unlike for ridge or lasso). However, significant progress has been made of late. We contribute to this literature by providing fundamental algebraic and statistical results for the minimum ℓ_2 -norm OLS interpolator. Whereas recent research primarily focuses on the OLS interpolator's prediction accuracy, our results primarily focus on parameter estimation, which is often the central object in statistics, especially in causal inference, as these parameters are endowed with causal meanings. In particular, we provide algebraic equivalents of (i) the leave-k-out formula, (ii) the omitted variable bias formula, and (iii) the Frisch-Waugh-Lovell theorem in the overparameterized regime. Under the Gauss-Markov model, we present statistical results such as an extension of the Gauss-Markov theorem and an analysis of variance estimation under homoskedasticity for the overparameterized regime.

3 - Towards a New Frontier of Trustworthy AI: Interpretable Machine Learning Algorithms that Produce All Good Models

Chudi Zhong, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Machine learning has been increasingly deployed for high-stakes decisions that deeply impact people's lives. It's crucial to ensure the safe and efficient utilization of machine learning models in the decision-making process. In this talk, I will introduce a new paradigm, called learning the Rashomon set, which finds and stores all models within epsilon of the optimal loss. I will present algorithms for finding optimal models and Rashomon sets, discuss how this new paradigm can break the interaction bottleneck between users and ML algorithms, and provide examples of its applications.

4 - Generalized Additive Modeling for Survival Analysis

Madeleine Udell, Stanford University, Stanford, CA, United States, Mike Van Ness, Ziran Zhou

Additive models offer accurate and interpretable predictions for tabular data, a critical tool for statistical modeling. Recent advances in Neural Additive Models (NAMs) allow these models to handle complex machine learning tasks, including feature selection and survival analysis, on large-scale data. This talk introduces dnamite, a Python package that implements NAMs for these advanced applications. dnamite provides a simple scikit-learn style interface to train regression, classification, and survival analysis NAMs, with built-in support for feature selection. We describe the methodology underlying dnamite, its design principles, and its implementation. Through an application to the MIMIC III clinical dataset, we demonstrate the utility of dnamite in a real-world setting where feature selection and survival analysis are both important.

5 - Mixed-Integer Quadratic Optimization with Switching Constraints

Andres Gomez, University of Southern California, Los Angeles, CA, United States, Shaoning Han, Soobin Choi, Valentina Cepeda

We investigate convexification in convex quadratic optimization with step function penalties. Such problems can be cast as mixed-integer quadratic optimization problems, where binary variables are used to encode the non-convex step function. First, we derive the convex hull for the epigraph of a quadratic function defined by a rank-one matrix. Using this rank-one convexification, we develop copositive and semi-definite relaxations for general convex quadratic functions. Leveraging these findings, we construct convex formulations to the support vector machine problem with 0-1 loss and show that they yield robust estimators in settings with anomalies and outliers.

TB35

Building B Level 2 B209

Humanitarian and Disaster Response Logistics

Invited Session

Public Sector OR

Chair: Alican Yilmaz, Northeastern University, Boston, United States

Co-Chair: Ozlem Ergun, Northeastern University, Newton, MA, United States

1 - Inventory Prepositioning as a Resilience Strategy Under Disruption

Alican Yilmaz, Northeastern University, Boston, MA, United States

The resilience of supply chain networks has become increasingly critical due to rising disruptions from natural disasters, trade conflicts, and terrorist activities. The severity of these disruptions highlights the need for effective resilience measures. In this study, we propose a tri-level defender-attacker-defender (DAD) model to enhance supply chain resilience through inventory prepositioning. The DAD problem is known for its high computational complexity, with exact algorithms in the literature often facing significant scalability challenges as problem sizes increase. To mitigate this issue, we adopt a network science based approach to identify candidate critical components, thereby reducing problem size and computational overhead. Specifically, we introduce a novel node criticality measure based on betweenness centrality to identify high-impact disruption targets. This measure is evaluated across various supply chain network configurations to assess its effectiveness. Next, we integrate this criticality information into the Column-and-Constraint Generation (C&CG) framework, reducing the number of iterations required for global optimality and thereby improving the runtime performance. Finally, we validate the proposed measure through a case study on the Alaska fuel supply chain network, offering insights into its effectiveness and identifying scenarios where it may be less effective.

2 - Multistage Chance-Constrained Programming

Soumya Ranjan Pathy, Clemson University, Clemson, SC, United States, Hamed Rahimian

In this talk, we explore multistage chance-constrained programming, a framework for sequential decision-making under uncertainty. This work introduces novel decomposition-based cutting-plane methodologies, specifically tailored to efficiently address these challenging optimization problems. Utilizing a ventilator allocation problem, we provide preliminary computational results that demonstrate the effectiveness and practical viability of the proposed approaches. Additionally, we offer managerial insights derived from evaluating multistage policies with varying decision frequencies over a fixed planning horizon. Our discussion underscores the critical impact of decision timing within a multistage optimization framework, emphasizing its relevance and potential strategic advantages in real-world decision-making contexts.

3 - Hurricane Shelter Networks Under Evacuation Behavior: Insights from Large-Scale Smartphone Data

Yu Lu, Tsinghua University, Beijing, China, People's Republic of, Fernanda Bravo, Elisa Long

Widespread natural disasters such as hurricanes and wildfires pose critical challenges for spatial resource allocation, including shelter site selection to support mass evacuations. Decision-makers must balance accessibility, coverage, and equity to ensure effective disaster response. We develop a 2-stage estimate-then-optimize model that integrates individual-level evacuation behavior into a shelter location optimization model. Using smartphone geolocation data from 1.57 million individuals during Hurricane Irma, we estimate a multinomial logit model (MNL) of evacuation choices---to a shelter, hotel, other location, or to stay home---accounting for demographics, shelter proximity, and evacuation orders.

An estimated 29% of Florida residents evacuated during Hurricane Irma, with most evacuees (87%) traveling to other locations, followed by shelters (8%) and hotels (6%). Evacuees are highly sensitive to distance, with stronger preferences for nearby shelters than hotels, and residents from mobile home communities and rural regions more frequently evacuate to shelters. Counterfactual analyses demonstrate that an Optimal-MNL policy increases predicted shelter evacuations by 28% and reduces the average shelter evacuation distance by 16%, relative to the status quo (445 shelters in operation in Florida). Compared to other benchmark policies, the Optimal-MNL policy delivers more equitable shelter access, particularly for suburban, rural, and mobile-home communities. This research offers actionable insights for disaster management, demonstrating that strategic shelter placement can achieve better outcomes with fewer facilities. Model robustness is validated through application to Hurricane Milton.

4 - Integrated Planning of Drone-Based Disaster Relief: Facility Location, Inventory Prepositioning, and Fleet Operations Under Uncertainty

Yuyang Ma, University of Southern California, Los Angeles, CA, United States, Karmel Shehadeh, Man Yiu Tsang

We introduce a two-stage robust optimization (RO) framework for the integrated planning of a drone-based disaster relief operations problem (DDROP). In the first stage, the model decides the number and locations of drone and relief items storage facilities, drone fleet size, drone assignment to demand nodes, and the amount of relief items to procure. In the second stage, the model decides drone reassignment and relief item distributions. We equip the RO model with a decision-dependent uncertainty set, which captures correlations between facility and drone disruptions and between demands and facility disruptions. We adopt the K -adaptability approximation scheme to address the combined challenges of integer recourse and decision-dependent uncertainty. We develop a new hybrid constraint-and-column generation (h-C&CG) algorithm to solve the K -adaptability approximation. Additionally, we derive new symmetry-breaking constraints and valid inequalities that improve the solvability of the model. We conduct extensive numerical experiments to demonstrate the computational efficiency of the proposed h-C&CG algorithm. Our results also show the superior operational performance of our proposed approach over traditional methods and illustrate the value of incorporating drone reassignments into disaster response planning.

5 - Improving Supply Chain Network Resilience: a Framework for Diagnosing Vulnerabilities in E-Commerce Distribution Networks

Jarrod Goentzel, Massachusetts Institute of Technology, Cambridge, MA, United States, Chelsey Graham, Ozlem Ergun

Supply chains have become more complex with interconnections and extended reach as customer expectations have risen. Simultaneously, supply chains designed for efficiency have increasingly had to adapt to disruptions such as extreme weather events, cyber security attacks, and a global pandemic. Therefore, companies now seek to consider resilience as part of their supply chain network design in order to remain competitive. In this paper, we investigate the vulnerability of e-commerce distribution networks through a three-step methodological approach: 1) network generation using optimization-based design principles, 2) network disruption, and 3) network evaluation. By studying many instances of regional networks in the U.S. and thousands of disruption scenarios, we explore which network structures best facilitate resilience and develop guidance for adaptation strategies.

TB36

Building B Level 2 B210

Recent Advances in Stochastic and Nonlinear Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Sanyou Mei, The Hong Kong University of Science and Technology, Hong Kong, N/A

1 - Revisiting Smoothing-Enabled Stochastic Gradient-Type Schemes**Uday Shanbhag, University of Michigan, Ann Arbor, University Park, PA, United States**

We revisit the classical randomized smoothing techniques employed with stochastic gradient-type schemes. Recent progress is discussed in a broad range of settings afflicted by nonconvexity and hierarchy.

2 - Developing Lagrangian-Based Methods for Nonsmooth Nonconvex Optimization**Kuangyu Ding, Purdue University, West Lafayette, IN, United States, Nachuan Xiao, Xiaoyin Hu, Kim-Chuan Toh**

In this paper, we consider the minimization of a nonsmooth nonconvex objective function $f(x)$ over a closed convex subset $\mathcal{X}$ of $\mathbb{R}^n$, with additional nonsmooth nonconvex constraints $c(x) = 0$. We develop a unified framework for developing Lagrangian-based methods, which takes a single-step update to the primal variables by some subgradient methods in each iteration. These subgradient methods are “embedded” into our framework, in the sense that they are incorporated as black-box updates to the primal variables. We prove that our proposed framework inherits the global convergence guarantees from these embedded subgradient methods under mild conditions. In addition, we show that our framework can be extended to solve constrained optimization problems with expectation constraints. Based on the proposed framework, we show that a wide range of existing stochastic subgradient methods, including the proximal SGD, proximal momentum SGD, and proximal ADAM, can be embedded into Lagrangian-based methods. Preliminary numerical experiments on deep learning tasks illustrate that our proposed framework yields efficient variants of Lagrangian-based methods with convergence guarantees for nonconvex nonsmooth constrained optimization problems.

3 - Accelerated Stochastic Approximation with State-Dependent Noise**Tianjiao Li, Georgia Institute of Technology, Atlanta, GA, United States, Sasila Ilandarideva, Anatoli Juditsky, Guanghui Lan**

We consider a class of stochastic smooth convex optimization problems under rather general assumptions on the noise in the stochastic gradient observation. As opposed to the classical problem setting in which the variance of noise is assumed to be uniformly bounded, herein we assume that the variance of stochastic gradients is related to the “sub-optimality” of the approximate solutions delivered by the algorithm. Such problems naturally arise in a variety of applications, in particular, in the well-known generalized linear regression problem in statistics. However, to the best of our knowledge, none of the existing stochastic approximation algorithms for solving this class of problems attain optimality in terms of the dependence on accuracy, problem parameters, and mini-batch size.

We discuss two non-Euclidean accelerated stochastic approximation routines—stochastic accelerated gradient descent (SAGD) and stochastic gradient extrapolation (SGE)—which carry a particular duality relationship. We show that both SAGD and SGE, under appropriate conditions, achieve the optimal convergence rate, attaining the optimal iteration and sample complexities simultaneously. However, corresponding assumptions for the SGE algorithm are more general; they allow, for instance, for efficient application of the SGE to statistical estimation problems under heavy tail noises and discontinuous score functions. We also discuss the application of the SGE to problems satisfying quadratic growth conditions, and show how it can be used to recover sparse solutions. Finally, we report on some simulation experiments to illustrate numerical performance of our proposed algorithms in high-dimensional settings.

4 - Asgo: Adaptive Structured Gradient Optimization**Kang An, Rice University, Houston, TX, United States**

Optimizing deep neural networks presents a unique challenge: their parameters are inherently structured as matrices and tensors, departing from traditional vector representations. This structure gives rise to critical, yet often underutilized, properties such as low-rank gradients and approximately block-wise diagonal Hessians. While these properties are key to developing efficient optimization algorithms, many widely used optimizers, including Adam, do not fully leverage them. I will introduce ASGO (Adaptive Structured Gradient Optimizer), a new algorithm that explicitly incorporates these structural insights. We achieve this by employing an adaptive preconditioner that responds directly to the structured gradients. I will present our fine-grained theoretical analysis demonstrating ASGO's superior convergence rates over existing structured gradient methods, and discuss how ASGO directly benefits from low-rank and block-wise diagonal properties. Finally, I will showcase ASGO's practical effectiveness with empirical results on challenging language model tasks.

5 - Nonasymptotic Analysis of Accelerated Methods with Inexact Oracle Under Absolute Error Bound**Yin Liu, Peking University, Beijing, China, People's Republic of, Sam Davanloo**

In many emerging applications, access to exact gradient information is either computationally expensive or impractical, giving rise to the need for optimization algorithms that are robust to inexact oracles. While accelerated first-order methods offer faster convergence in the exact setting, they tend to be more sensitive to gradient errors compared to their non-accelerated counterparts.

This work presents a nonasymptotic convergence analysis for two accelerated methods—Inexact Fast Gradient Method (FGM) and Inexact Optimized Gradient Method (OGM)—for solving deterministic smooth convex optimization problems under inexact gradient information. Our analysis employs the Performance Estimation Problem (PEP) framework to derive analytical convergence bounds for these algorithms when the gradient approximation errors vary over iterations.

Under an absolute error model, we obtain the bounds where the cumulative error term is shown to be independent of initial conditions. We further investigate the trade-off between convergence rate and error accumulation, leading to insights on selecting optimal step sizes. In addition, we characterize the optimal scheduling of gradient inexactness across iterations to ensure that error accumulation remains subordinate to the convergence rate, thereby preserving the fast convergence of the algorithm.

TB37

Building B Level 2 B211

Recent Advances in Data-Driven Optimization

Invited Session

OPT: Optimization Under Uncertainty

Chair: Soroosh Shafiee, Cornell University, Ithaca, NY, United States

Co-Chair: Sloan Nietert, Cornell University, Ithaca, NY, United States

1 - Estimation of Stochastic Optimal Transport Maps**Sloan Nietert, Cornell University, Ithaca, NY, United States**

The optimal transport (OT) map is a geometry-driven transformation between high-dimensional probability distributions which underpins a wide range of tasks in statistics, applied probability, and machine learning. However, existing statistical theory for OT map estimation is quite restricted, hinging on Brenier's theorem (quadratic cost, absolutely continuous source) to guarantee existence and uniqueness of a deterministic OT map, on which various additional regularity assumptions are imposed to obtain quantitative error bounds. In many real-world problems these conditions fail or cannot be certified, in which case optimal transportation is possible only via

stochastic maps that can split mass. To broaden the scope of map estimation theory to such settings, this work introduces a novel metric for evaluating the transportation quality of stochastic maps. Under this metric, we develop computationally efficient map estimators with near-optimal finite-sample risk bounds, subject to easy-to-verify minimal assumptions. Our analysis further accommodates common forms of adversarial sample contamination, yielding estimators with robust estimation guarantees. Empirical experiments are provided which validate our theory and demonstrate the utility of the proposed framework in settings where existing theory fails. These contributions constitute the first general-purpose theory for map estimation, compatible with a wide spectrum of real-world applications where optimal transport may be intrinsically stochastic.

2 - Data-Driven Robust Inventory Planning and Advertising Under Product Return

Qinshen Tang, Nanyang Technological University, Singapore, Singapore, Hongyan Zhang, Yugang Yu, Ye Shi, Li Chen

Retailers increasingly use online pay-per-click (PPC) ads on platforms like Amazon to boost sales, but these ads also influence product returns—an effect largely overlooked in prior research. Motivated by a real-world case, we study a multi-product setting where online ads impact both sales and returns, complicating inventory and advertising planning. We conduct predictive analysis on historical data and identify both direct effects (a product's ads influencing its own sales and returns) and spillover effects (ads affecting other products). Based on this, we develop linear models to estimate net demand and returns. To address data limitations and distributional uncertainty, we propose a robust satisficing framework that allows the retailer to set a desired profit level and make data-driven ad and inventory decisions accordingly. A case study demonstrates the practical scalability and improved performance of our Linear Approximation Robust Satisficing (LARS) model over the traditional Predict-then-Optimize (PO) approach. The LARS model achieves superior results, especially when spillover effects are present, outperforming PO by an average of 3.82%.

3 - Scaling the Cvar Approximation for Solving Chance Constrained Programs

Nan Jiang, Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Rui Chen

In a chance constrained program (CCP), the decision maker aims to find an optimal solution that ensures the probability of violating uncertainty constraints remains within a specified risk level. However, the nonconvex nature of the CCP poses significant challenges in computing an optimal solution. To address this, convex approximations have been developed, with Conditional Value-at-Risk (CVaR) recognized as one of the best practical methods. Despite its strengths, CVaR approximation can be overly conservative. In this paper, we propose a scaling approach that enhances the CVaR approximation. We identify sufficient conditions under which the scaled CVaR approximation preserves optimality and demonstrate its applicability to both single and joint chance constrained programs. Since solving the scaled CVaR approximation to optimality is NP-hard, we develop efficient algorithms to iteratively refine the scaling factors and approximately solve the scaled CVaR approximation. Numerical experiments validate the proposed methods, confirming improvements over the CVaR approximation.

4 - Distributionally Robust Optimization via Iterative Algorithms in Continuous Probability Spaces

Linglingzhi Zhu, Georgia Institute of Technology, Atlanta, GA, United States, Yunqin Zhu, Yao Xie

We consider a minimax problem motivated by distributionally robust optimization (DRO) when the worst-case distribution is continuous, leading to significant computational challenges due to the infinite-dimensional nature of the optimization problem. Recent research has explored learning the worst-case distribution using neural network-based generative models to address these computational challenges but lacks algorithmic convergence guarantees. This paper bridges this theoretical gap by presenting an iterative algorithm to solve such a minimax problem, achieving global convergence under mild assumptions. In particular, we reformulate the discrepancy-based DRO as a minimax problem in the Wasserstein space and, by applying Brenier's theorem, represent the worst-case distribution as a transport map from a continuous reference measure for computation. Furthermore, we demonstrate that the worst-case distribution can be efficiently computed using the Jordan-Kinderlehrer-Otto (JKO) scheme with sufficiently large regularization parameters for commonly used discrepancy functions, linked to the radius of the ambiguity set. Additionally, we derive the global convergence rate and quantify the total number of subgradients and inexact JKO iterations required to obtain approximate stationary points. Numerical results on an illustrative robust classification example demonstrate the effectiveness of our DRO framework. Compared to existing point-to-point robust training methods, the solution derived from the proposed algorithm serves as an efficient adversarial attack generator thanks to the continuous probability space formulation.

5 - Learning Optimal Inventory Control Policies with Deep Neural Nets

Zhen Yang, University of Texas at Austin, Austin, TX, United States, Yizhe Huang, Rui Gao, Shuang Li

The inventory management problem is crucial for many businesses and has been extensively studied in previous works. Despite these efforts, some models remain challenging due to the unknown structure of the optimal inventory policy. We aim to approximate the policy at each stage by a neural network which is a universal approximator, and establish the statistical and global optimality guarantees.

TB38

Building B Level 2 B212

Recent Progress in Optimization Software II

Invited Session

OPT: Computational Optimization and Software

Chair: Hans Mittelmann, Arizona State University, Tempe, AZ, United States

1 - Recent Developments in Optverse Optimization Solver

Zirui Zhou, Huawei Technologies Canada, Burnaby, BC, Canada

In this talk, we present latest development of Optverse optimization solver and some of its successful applications. We will introduce new features, new solvers, and discuss some techniques that are key to performance improvement.

2 - What's New in FICO Xpress Solver

Michael Perregaard, FICO, Birmingham, United Kingdom

We will give an overview of the latest enhancements, the newest features, and the most recent performance improvements in the FICO Xpress Solver for mixed-integer linear and nonlinear optimization problems. These include advances in MIP techniques such as heuristics, cutting, branching and presolving.

3 - Highs: Latest Updates

Benjamin Champion, University of Edinburgh, Edinburgh, United Kingdom

HiGHS is the world's leading open-source linear optimization solver. It supports Linux, macOS, and Windows, and interfaces with many popular programming languages. This talk examines recent developments and future prospects, primarily focusing on technical and algorithmic aspects. However, this talk also discusses other important parts of the HiGHS project. Fundamentally, a thriving software ecosystem stimulates development. The HiGHS team interacts with a range of users and contributors, a substantial proportion of whom are industrial. These interactions take many forms, including usage queries, feature requests, code contributions, and sponsorship. This talk outlines the key proceedings at the second annual HiGHS Workshop, an event bringing together many members of the HiGHS ecosystem. As with any open-source software project, planning for long-term survival is crucial. This talk presents some thoughts on the evolution of the HiGHS team.

4 - Latest Developments in the COPT Optimizer

Joachim Dahl, Cardinal Operations, Ballerup, Denmark

We discuss recent advances in the COPT optimizer from Cardinal Operations. The advances include improvements in the mixed-integer optimizer for various problem classes, as well as improvements of the conic optimizer.

5 - Recent Developments in the Gurobi Optimizer

Dan Steffy, Gurobi Optimization, Bloomfield Hills, MI, United States

This talk will discuss new features and performance enhancements in the Gurobi Optimizer for LP, MILP and Global MINLP. We will also present preliminary performance results for the upcoming release.

TB39

Building B Level 2 B213

Theoretical and Computational Advances in Branch-and-Cut Algorithms

Invited Session

OPT: Integer and Discrete Optimization

Chair: Aleksandr Kazachkov, University of Florida, Gainesville, United States

1 - Generalization Guarantees for Learning Branch-and-Cut Policies

Hongyu Cheng, Johns Hopkins University, Baltimore, MD, United States, Amitabh Basu

Mixed-integer programming (MIP) provides a powerful framework for optimization problems, with Branch-and-Cut (B&C) being the predominant algorithm in state-of-the-art solvers. The efficiency of B&C critically depends on heuristic policies for making sequential decisions, including node selection, cut selection, and branching variable selection. While traditional solvers often employ heuristics with manually tuned parameters, recent approaches increasingly leverage machine learning, especially neural networks, to learn these policies directly from data. A key challenge is to understand the theoretical underpinnings of these learned policies, particularly their generalization performance from finite data. This paper establishes rigorous sample complexity bounds for learning B&C policies where the scoring functions guiding each decision step (node, cut, branch) have a certain piecewise polynomial structure. This structure generalizes the linear models that form the most commonly deployed policies in practice and investigated recently in a foundational series of theoretical works by Balcan et al. Such piecewise polynomial policies also cover the neural network architectures (e.g., using ReLU activations) that have been the focal point of contemporary practical studies. Consequently, our theoretical framework closely reflects the models utilized by practitioners investigating machine learning within B&C, offering a unifying perspective relevant to both established theory and modern empirical research in this area. Furthermore, our theory applies to quite general sequential decision making problems beyond B&C.

2 - Extended Abstract Branch-and-Cut Model

Boyang Han, University of Florida, Gainesville, FL, United States, Aleksandr Kazachkov

We extend the abstract branch-and-cut tree frameworks considered by Kazachkov, Le Bodic, and Sankaranarayanan [3], in which branching leads to known (constant) bound improvements while cut nodes provide diminishing strength across rounds and potentially increase node solving time. We generalize these models to optimize branching and cutting decisions to minimizing solving time when incorporating cut generation time and multiple cut families in the same model as both diminishing-strength cuts and general monotonic node-processing time functions. We prove that it is sufficient to only add root cuts in the symmetric branching case (when branching in each direction provides the same bound improvement). Next, we analyze optimal tree structure in the nonsymmetric branching setting, for which we show that local cuts (beyond the root node) can be necessary in optimal trees.

3 - Speeding up Mixed-Integer Solvers with Sparse Learning for Branching

Selin Bayramoglu, Georgia Institute of Technology, Atlanta, GA, United States, George Nemhauser, Nikolaos Sahinidis

Machine learning is being increasingly used to accelerate branch-and-bound algorithms for mixed integer optimization. Studies in the literature demonstrate success in approximating branching criteria based on strong branching, a local and expensive strategy which empirically leads to small search trees. The state-of-the-art learning methods for this problem are based on neural networks, with many parameters and a need for a large amount of data for training. In this work, we demonstrate the effectiveness of sparse models in approximating strong branching scores, showing that our models remain simple yet powerful even when trained on limited data. Moreover, our approach achieves significant solver speedups without requiring GPU hardware. We demonstrate that our methodology speeds up SCIP when applied to commonly studied mixed-integer linear problems, and a challenging mixed-integer nonlinear problem from the energy sector.

4 - Learning to Parallelize Branch & Bound

Kavitha G Menon, Argonne National Laboratory, Lemont, IL, United States, Santanu Dey, Alinson Santos Xavier

Branch-and-bound is theoretically well suited for parallel processing, but achieving significant speedups in practice is challenging due to imbalanced workloads and communication overhead. In this study, we propose a new learning-based parallelization approach, which statically partitions Mixed-Integer Programming (MIP) problems into well-balanced subproblems, avoiding both issues. We first solve a representative set of MIP instances with MIPLearn, an open-source package for learning-enhanced optimization, capture the B&B tree, then identify B&B tree edges whose removal creates balanced subtrees. This data is then used to train machine learning models capable of partitioning new MIP instances. Unlike existing methods, our approach allows edge removal throughout the entire tree, not just near the root, resulting in more balanced partitions. Computational results, conducted on eight families of discrete combinatorial problems, show that the method is able to significantly outperform Gurobi's parallel performance, often achieving near-linear speedups up to 16 processors. The study demonstrates the effectiveness of parallelizing B&B through strategic partitioning and highlights areas for further optimization.

5 - Recent Developments in the Gurobi MILP Solver

Ed Klotz, Gurobi Optimization, Incline Village, NV, United States

Performance improvements to the Gurobi MILP Solver will be discussed. Topics include extending large neighborhood search heuristics such as RINS (Relaxation Induced Neighborhood Search) to make use of infeasible integral vectors, and extensions to Optimization Based Bound Tightening (OBBT) to handle specific constructs common to some challenging MILP and MINLP problems.

TB40

Building B Level 2 B214

Platform and Service Operations

Contributed Session

Contributed

Chair: Saman Modiri, Lehigh University, Emmaus, United States

1 - Network Formation Through Mechanism Design

Bo Cowgill, Columbia University, New York, NY, United States, Zikai Xu

This paper develops a mechanism design approach to network formation. A principal has a willingness-to-pay (WTP) for different network configurations, while agents have preferences over their network positions. Our approach allows the principal to optimize for global properties of the network, while respecting IC/IR constraints of network participants. We focus on direct mechanisms, but develop a broader family of mechanisms in which transfers are set by the allocation rule ("revenue equivalence"). We characterize optimal mechanisms under novel multidimensional regularity conditions, and provide an ironing procedure for irregular distributions. These findings contribute to multidimensional mechanism design, with potential applications to network formation in social, economic, and organizational contexts.

2 - Incentivizing In-Store Customers for Last-Mile Delivery: a Dynamic Compensation Framework

Rouven Schur, University Duisburg-Essen, Duisburg, Germany, Matthias Soppert, Kai Winheller

Last-mile delivery remains one of the most challenging and cost-intensive components of modern retail logistics. A promising approach to alleviate these challenges is to crowdsource deliveries to in-store customers who, after completing their own shopping, may be willing to transport goods along their homeward routes. These occasional drivers (ODs) can serve as a flexible and cost-efficient supplement to traditional delivery fleets. However, their successful integration into delivery operations hinges on overcoming two challenges: the uncertainty of ODs' destinations and their voluntary decision to accept delivery tasks.

We address these challenges through the application of dynamic pricing, using monetary incentives to influence OD behavior in real time. The retailer's decision problem is modeled as a finite-horizon Markov decision process, where compensation offers are dynamically adapted to the current system state – specifically, the set of remaining delivery locations and the time left to attract ODs before tasks must be fulfilled by dedicated drivers. In contrast to prior work, our model does not assume knowledge of ODs' destinations or guaranteed participation. Instead, it captures the stochasticity in both OD availability and willingness to accept tasks.

We derive analytical results for a one-dimensional setting and identify structural properties of the optimal policy. These insights enable the development of a decomposition-based heuristic for the more complex two-dimensional case. A simulation study demonstrates the benefits of a dynamic, state-dependent compensation scheme in effectively engaging in-store customers for last-mile delivery.

3 - Dynamic Control of Add-on Services in Queueing Systems

Yuan-Mao Kao, Baruch College, CUNY, New York City, NY, United States, Chen-An Lin

We study a dynamic control problem in a service system where a provider offers an add-on service to each customer after completing a primary service. Customers choose whether to use the primary and add-on services based on their evaluations of service quality, prices, and expected wait times. Their expectations for the quality of the add-on service depend on their experiences with the primary service. The provider can dynamically adjust prices and service rates to manage congestion and customer experience in order to optimize long-run performance.

4 - Pricing and Wage Decisions for on-Demand Food Delivery Platforms with Multiple Customer Classes and Courier Pools

Yooseok Seo, Seoul National University, Seoul, Korea, Republic of, Jungwook Lim, Byeongkwon Lee, Kunsoo Park

The recent proliferation of food delivery services with crowdsourced couriers (delivery workers) has introduced new problems in the operations of on-demand food delivery platforms. An increase in crowdsourced couriers has diversified delivery transportation modes such as motorcycle, bike, and walk, which results in mode-specific maximum delivery distances according to their speeds. In addition, it is necessary to manage various customer classes with specific delivery distances between restaurants and customers.

For a food delivery platform with multiple customer classes and courier pools, it is essential to optimize the delivery price charged to each customer, the wage paid to couriers, and matching between them to complete the food delivery within a reasonable amount of time. Hence, we study the operations decisions for the online food delivery platform that faces price- and time-sensitive customers with heterogeneous delivery distances and earning-sensitive couriers with heterogeneous transportation speeds. We formulate this problem as a multi-class multi-pool queueing and allocation problem.

We find that the objective function of this optimization problem is inherently complex with respect to the number of customer classes and courier pools, making it unsuitable for standard convex optimization techniques.

Hence, we propose a new Utilization-Based Multipoint (UBM) heuristic that exploits multiple initial points to derive near-optimal results efficiently. Our numerical experiments on both generated instances and real delivery data demonstrate that the proposed heuristic algorithm significantly outperforms the benchmark for all experiments.

5 - Coach Reservation for Groups Requests

Carlos Cardonha, University of Connecticut, Storrs, CT, United States, Arvind Raghunathan

Passenger transportation is a cornerstone of a railway company's business, with ticket sales playing a pivotal role in revenue generation. This work extends traditional revenue management models by addressing the coach reservation problem for group requests, where two or more passengers must be seated in the same coach. We propose an exact mathematical programming formulation incorporating a first-come, first-served fairness condition for the offline case, in which all requests are known in advance. We also propose algorithms for online models of the problem. Our analysis for one of these models overcomes known barriers in the packing literature, yielding strong competitive ratio guarantees when group sizes are relatively small compared to coach capacity. Using data from the Shinkansen Tokyo-Shin-Osaka line, our numerical experiments demonstrate the practical effectiveness of the proposed policies.

6 - Analysis of the Revenue Sharing Ratio of A Virtual Product Between A Developer and A Distribution Platform Under A Consignment Contract with Revenue Sharing

SEONG-HYUN NAM, University of North Dakota, GRAND FORKS, ND, United States

Virtual products such as software and applications, movie and music streaming, online games, e-books, and mobile applications have been rapidly increasing their market share recently. We consider the case where owners or developers of virtual products contract to sell their virtual products through a third-party sales or distribution platform. Different types of contracts can be established between the owner (developer) of a virtual product and the distribution platform. In this study, we considered in particular a consignment contract with revenue sharing. In such a contract, the developer transfers ownership of the virtual product and sets its selling price, with the platform charging the developer an agreed percentage of the selling price for each virtual product sold. Under a consignment contract with revenue sharing, we examined the range of revenue sharing percentages that developers can accept when they delegate everything to the sales platform, how the revenue sharing percentage differs when developers use a hybrid format in which they sell directly through their own sales channels (direct channel) and the sales platform (indirect channel) compared to when they delegate everything to the sales platform, and how much the accuracy of developer and platform demand forecasts affects the revenue sharing percentage.

7 - Political Posturing: Market Expectations, Profit Maximization, and Consumer Surplus

Saman Modiri, Lehigh University, Emmaus, PA, United States

Brand activism has become an increasingly prominent aspect of firm strategy, particularly as companies navigate heterogeneous consumer segments. However, firms have demonstrated inconsistent patterns—shifting between political engagement, neutrality, and even ideological reversal—raising questions about the strategic logic underlying such actions. This study develops a stylized analytical model in which firms select their political alignment to maximize profit, accounting for consumer heterogeneity, issue salience, and market expectations. The model further incorporates consumer surplus to evaluate the broader implications of political posturing on market outcomes. Findings reveal that when political alignment is expected by the market, firms gain from activist positioning; yet, consumer surplus declines as the issue becomes more salient in daily life. In contrast, for issues with lower salience, firms can satisfy market expectations with minimal surplus distortion and, at times, even an increase in consumer surplus, leading to potential gains for both the firm and consumers. Additionally, when neutrality is not penalized, firms face fewer pressures to signal alignment, and consumer surplus improves. These results highlight brand activism as a strategic tool driven by market expectations, consumer preferences, and the perceived relevance of political issues. This research provides a theoretical foundation for interpreting recent shifts in corporate political behavior and contributes to a broader understanding of the market dynamics underpinning brand activism.

TB41

Building B Level 3 B301

Operational Perspectives to Clinical Trial Design

Invited Session

Health Applications Society

Chair: Amin Khademi, Clemson University, Clemson, SC, United States

1 - To Abandon OR Not? Understanding the Impact of Participant Retention on Clinical Trial Outcomes

Xueze Song, The University of Alabama, TUSCALOOSA, AL, United States, Mili Mehrotra, Ujjal Mukherjee, Tharanga Rajapakshe

The process of drug development is replete with significant risks and uncertainties, with as many as 12% of clinical trials terminated early. One leading factor of trial termination is poor recruitment or retention. In this work, we study the impact of strategies to improve participant retention on clinical trial sponsor's decision to terminate the clinical trial. Specifically, we address the following research questions: How does the adoption of incentives and effort affect the sponsor's decision to terminate the clinical trial? How does the impact of such strategies vary across different clinical trials? How can the sponsor adjust their retention strategies to minimize the likelihood of termination? To answer these questions, we formulate a sponsor's decision to terminate a clinical trial as an optimal stopping problem with partial information on the drug's efficacy. We then leverage the data from the clinical study registry from the U.S. and empirically estimate the effect of various clinical trial characteristics on the likelihood of termination. Our analysis sheds light on the key factors that drive clinical trial termination and can inform future drug development efforts.

2 - Optimal Enrollment in Late Stage New Drug Development with Learning of Drug's Efficacy for Group Sequential Clinical Trials

Zhili Tian, Univ. of Houston, Houston, TX, United States, Gordon Hazen, Hong Li

The cost for developing a new drug ranges is close to more than \$2 billion between 2010 and 2019. The efficacy of the candidate drug, patient enrollment, the market exclusivity period (MEP), and the planning horizon are uncertain. Slow enrollment leads to increased costs, canceled clinical trials, and lost potential revenue. Firms, hoping to detect efficacy versus futility of the candidate drug early to save development costs, plan interim analyses (IAs) of patient response data.

We developed a continuous-time dynamic programming (DP) model with learning of a drug's efficacy to assist firms in developing optimal enrollment policies in their clinical trials. We also established the optimality equation for this DP model. Through a clinical trial for testing a cancer drug developed by a leading pharmaceutical firm, we demonstrate that our DP model can help firms effectively manage their trials with a sizable profit gain. Firms can also use our model in simulation to select their trial design parameters (e.g., the sample sizes of IAs).

We established that the value of the drug development project increases as the average response from patients using the candidate drug increases. For drugs having low annual revenue or a strong market brand or treating rare diseases, we also established that the optimal enrollment policy is monotonic in the average patient response. The optimal enrollment rate increases as the variance of the MEP decreases.

Firms can use our optimal policy to guide patient recruitment in clinical trials considering competition from other drugs in the market place.

3 - Online Learning with Survival Data

Arielle Anderer, Cornell University, Johnson School, Ithaca, NY, United States, Hamsa Bastani, John Silberholz

In many healthcare settings it is imperative to be able to quickly converge to the best intervention, but the structure of the data makes using traditional methods challenging. This project adapts online learning techniques to scenarios with time-to-event data, such as time-to-screening and time-to-disease-progression, where there is a delay between choosing an arm and observing feedback that is endogenous to the quality of the arm. We posit a multi-armed bandit algorithm that uses a cox-proportional hazards estimator. We theoretically analyze and prove guarantees on the regret under this algorithm. Lastly, we examine its performance on a dataset of behavioral health interventions for cervical cancer screening, and compare it to that of other adaptive allocation schemes.

4 - Planning Micro-Randomized and Adaptive Trials with Nonstationary Multi-Armed Bandits

Yonatan Mintz, University of Wisconsin Madison, Madison, WI, United States, Fengxu Li, Qinyang He

Mobile health interventions enable personalized treatments in areas like diabetes, mental health, and weight loss. These interventions generally use reinforcement learning methods and multi armed bandits to personalize treatment recommendations to trial participants and improve resource allocation. However there are several key challenges that arise when using these methods. First, since the goal of the trial is to use the data for post hoc analysis it is critical to have sufficient statistical power when allocating treatments to participants. When treatments are personalized however, there is a risk that not all treatments will be dispensed sufficiently often to ensure high statistical power. Furthermore, if interventions given to participants are costly or impact participant engagement, it is critical that the randomization methods account for not just the maximum effect but also how much cost is incurred by the participants. In this talk we will discuss a specific class of non-stationary bandit models that apply in these scenarios, and examine how they can be adapted to meet the unique needs of clinical trial design. We will discuss both theoretical and computational results of these methods including regret and power guarantees. We will also present results from case studies in the areas of weight loss and physical activity and show how these methods can be applied in practice.

TB42

Building B Level 3 B302

Data-driven healthcare analytics

Invited Session

Health Applications Society

Chair: Mohammad Zhalechian, Indiana University, Kelley School of Business, Bloomington, IN, United States

1 - Early Warning Index for Patient Deteriorations in Hospitals

Yu Ma, Massachusetts Institute of Technology, Cambridge, MA, United States, Dimitris Bertsimas, kimberly villalobos

Capacity-constrained hospitals lack automated systems to harness the growing volume of heterogeneous clinical and operational data—ranging from vital signs and lab results to scheduling and patient flow metrics—to effectively forecast critical events. Early identification of patients at risk for deterioration is essential not only for patient care quality monitoring but also for physician care management. However, translating varied data streams into accurate and interpretable risk assessments poses significant challenges due to inconsistent data formats.

We develop a multimodal machine learning framework, the Early Warning Index, to predict the aggregate risk of ICU admission, emergency response team dispatch, and mortality. Using a dataset of patients at a large U.S. hospital, our approach automatically extracts features from both structured and unstructured electronic health record data. Key to EWI's design is a human-in-the-loop process: clinicians help determine alert thresholds and interpret model outputs, which are enhanced by explainable outputs to highlight clinical and operational factors driving each patient's risk. We deploy EWI in a hospital dashboard that stratifies patients into three risk tiers. It is currently used as a triage tool for proactively managing at-risk patients.

The proposed approach saves physicians valuable time by automatically sorting patients of varying risk levels, allowing them to concentrate on patient care. By further pinpointing specific risk drivers, the proposed model provides data-informed adjustments to care giver scheduling and allocation of critical resources. As a result, clinicians and administrators can avert downstream complications including costly procedures or high readmission rates and improve overall patient flow.

2 - When Less is More: Optimizing Prescription Alerts Under Fatigue

Hossein Piri, Haskayen School of Business-University of Calgary, Calgary, AB, Canada, Michael Lingzhi Li

As healthcare systems grow more complex, ensuring safe medication practices remains a critical challenge. Pharmacists play a key role in preventing medication errors, aided by Computerized Provider Order Entry (CPOE) systems that generate alerts for potential drug interactions, dosages, and contraindications. However, to capture all potential risks, these systems generate a significant number of alerts, leading to alert fatigue and diminishing pharmacists' responsiveness to critical warnings. This study develops a fluid model to optimize CPOE alert strategies while accounting for pharmacist fatigue. The key challenge lies in the interdependence of decisions: each alert shown increases fatigue, which in turn reduces the pharmacist's capacity to act on future alerts, creating a dynamic feedback loop. We derive the general optimal policy, which follows a decreasing threshold structure—alerts are initially shown selectively but become more frequent as fatigue accumulates. Our analysis further shows that the heavier the tails of the risk distribution, the more conservatively alerts should be shown. While the dynamic thresholding strategy is optimal, its implementation is challenging due to operational complexity and the potential for strategic behavior. We show that a constant threshold policy is asymptotically optimal for a broad class of risk distributions, offering a simpler alternative with strong performance. Numerical experiments across a wide range of scenarios demonstrate that the optimized alert strategy reduces patient risk by 8–9% while cutting the number of alerts by 40–50%, highlighting its robustness and practical impact.

3 - Mosaic: Model-Agnostic Optimal Subgroup Identification with Multi-Constraint for Improved Reliability

Wenxin Chen, Cornell University, New York, NY, United States, Weishen Pan, Kyra Gan, Fei Wang

Identifying subgroups that benefit from specific treatments using observational data is a critical challenge in personalized medicine. Most existing approaches solely focus on identifying a subgroup with an improved treatment effect. However, practical considerations, such as ensuring a minimum subgroup size for representativeness or achieving sufficient confounder balance for reliability, are also important for making findings clinically meaningful and actionable. While some studies address these constraints individually, none offer a unified approach to handle them simultaneously. To bridge this gap, we propose a model-agnostic framework for optimal subgroup identification under multiple constraints. We reformulate this combinatorial problem as an unconstrained min-max optimization problem with novel modifications and solve it by a gradient descent ascent algorithm. We further prove its convergence to a feasible and locally optimal solution. Our method is stable and highly flexible, supporting various models and techniques for estimating and optimizing treatment effectiveness with observational data. Extensive experiments on both synthetic and real-world datasets demonstrate its effectiveness in identifying subgroups that satisfy multiple constraints, achieving higher treatment effects and better confounder balancing results across different group sizes.

4 - A Responsible Framework for Assessing, Selecting, and Explaining Machine Learning Models in Cardiovascular Disease Outcomes Among People with Type 2 Diabetes: Methodology and Validation Study

Yang Yang, Georgia Institute of Technology, Atlanta, GA, United States

Building machine learning models that are interpretable, explainable, and fair is critical for their trustworthiness in clinical practice. Interpretability, which refers to how easily a human can comprehend the mechanism by which a model makes predictions, is often seen as primary consideration when adopting a machine learning model in health care. However, interpretability alone does not necessarily guarantee explainability, which offers stakeholders insights into a model's predicted outputs. Moreover, many existing frameworks for model evaluation focus primarily on maximizing predictive accuracy, overlooking the broader need for interpretability, fairness, and explainability. This study proposes a 3-stage machine learning framework for responsible model development through model assessment, selection, and explanation. We demonstrate the application of this framework for predicting cardiovascular disease (CVD) outcomes, specifically myocardial infarction (MI) and stroke, among people with type 2 diabetes (T2D). We developed several interpretable machine learning models (linear, tree-based, and ensemble) to predict the risks of MI and stroke among patients with diabetes. Our 3-stage framework first assesses these models via predictive accuracy and fairness metrics. Then, in the model selection stage, we quantify the trade-off between accuracy and fairness using area under the curve (AUC) and Relative Parity of Performance Scores (RPPS), wherein RPPS measures the greatest deviation of all subpopulations compared to the population-wide AUC. Finally, we quantify the explainability of the chosen models using methods such as SHAP (SHapley Additive exPlanations) and partial dependence plots to investigate the relationship between features and model outputs.

TB43

Building B Level 3 B303

Digital Innovation and AI

Invited Session

MSOM: iForm

Chair: Yao Cui, Cornell University, Ithaca, NY, United States

Co-Chair: Jingchen Liu, Renmin University of China, Beijing, N/A

1 - Nft Market Design: Resale Royalty, Decentralization, and Interoperability

Jingchen Liu, Renmin University of China, Beijing, China, People's Republic of, Yao Cui

The Non-Fungible Token (NFT) market is fast-growing, but increasing competition among platforms poses significant challenges in choosing the optimal operational strategies and finding the right ways to collaborate. This paper provides a theoretical investigation into NFT market design by focusing on two key platform policies: (1) royalty payments the original NFT creators receive when a resale transaction occurs, and (2) NFT interoperability, i.e., whether resale is allowed on a different platform than where the NFT was originally minted. We show that when platforms maintain centralized control over royalty policies and do not permit interoperability, the NFT equilibrium can take one of three forms depending on the degree of market speculation: "free royalty", where creators earn exclusively from the primary price; "free airdrop", where creators earn exclusively from the resale royalty; and "intermediate royalty", where creators earn from both sources. However, the equilibrium shifts when interoperability is permitted (always to free royalty) or when royalty decisions are decentralized to individual creators (always to free airdrop). Comparing the NFT equilibria under different policies, we find that under centralized royalty policies, NFT interoperability can create a win-win outcome by improving both platform profits and social welfare. In contrast, under decentralized royalty policies, NFT interoperability consistently results in a lose-lose outcome. These findings suggest that platforms should only collaborate on NFT interoperability if they maintain decision-making authority over royalty policies. In this case, we further identify that the win-win outcome is most likely to occur when market speculation is at moderately low levels.

2 - Generative AI and Organizational Structure in the Knowledge Economy

Fasheng Xu, University of Connecticut, Stamford, CT, United States, Jing Hou, Wei Chen, Karen Xie

The adoption of Generative Artificial Intelligence (GenAI) is fundamentally reshaping organizations in the knowledge economy. GenAI can significantly enhance workers' problem-solving abilities and productivity, yet it also presents a major reliability challenge: hallucinations, or errors presented as plausible outputs. This study develops a theoretical model to examine GenAI's impact on organizational structure and the role of human-in-the-loop oversight. Our findings indicate that successful GenAI adoption hinges primarily on maintaining hallucination rates below a critical level. After adoption, as GenAI advances in capability or reliability, organizations optimize their workforce by reducing worker knowledge requirements while preserving operational effectiveness through GenAI augmentation—a phenomenon known as deskilling. Unexpectedly, enhanced capability or reliability of GenAI may actually narrow the span of control, increasing the demand for managers rather than flattening organizational hierarchies. To effectively mitigate hallucination risks, many firms implement human-in-the-loop validation, where managers review GenAI-enhanced outputs before implementation. While the validation increases managerial workload, it can, surprisingly, expand the span of control, reducing the number of managers needed. Furthermore, human-in-the-loop validation influences GenAI adoption differently based on validation costs and hallucination rates, deterring adoption in low-error, high-cost scenarios, while promoting it in high-error, low-cost cases. Finally, productivity improvements from GenAI yield distinctive organizational shifts: as productivity increases, firms tend to employ fewer but more knowledgeable workers, gradually expanding managerial spans of control. Our research directly addresses calls for theoretical frameworks to understand how GenAI technologies reshape organizational structures and the future of work, while providing practical guidance for organizations navigating this transformation.

3 - Supporting Teachers to Support Students in Individualized Learning

Owen Wu, Kelley School of Business, Indiana University, Bloomington, IN, United States

Individualized learning platforms are designed to optimize student progress by adapting content to each learner's needs. However, these systems often assume that students can work independently once a learning path is established. In practice, many students encounter difficulties that require timely support from teachers—support that current systems do not effectively facilitate. At present, teachers must manually review student records to infer when intervention is needed, a process that is both time-consuming and error-prone. In collaboration with an individualized learning platform, we aim to develop an intelligent system that detects when students need help during individualized learning. Our model identifies signs of struggle or frustration and alerts teachers in real time. The system not only indicates when intervention is needed, but also provides insights into what kind of support may be most effective. By supporting teachers in their ability to support students, we strengthen the human element of individualized learning and move toward a more responsive and inclusive educational model.

4 - Economics of NFTs: The Value of Creator Royalties

Niuniu Zhang, University of California, Los Angeles, Los Angeles, CA, United States, Brett Falk, Bin Gu, Gerry Tsoukalas

Non-Fungible Tokens (NFTs) are transforming how content creators, such as artists, price and sell their work. A key feature of NFTs is the inclusion of royalties, which grant creators a share of all future resale proceeds. Although widely used, critics argue that sophisticated speculators, which dominate the NFT market, simply price in royalties upfront, neutralizing their impact. We show this intuition to be true only under the assumption of perfect frictionless markets. Under more realistic market conditions, royalties enable creators to capitalize on the presence of speculators in at least three ways: They can enable risk sharing (under risk aversion), mitigate information asymmetry (when speculators are better informed) and unlock price discrimination benefits (in multi-unit settings). Moreover, in all three cases, royalties also meaningfully expand trade, implying increased transaction volume for platforms. These theoretical results offer testable predictions that could drive future NFT economics research.

5 - Optimal Collateral for Clearing Interbank Payments

Shuzhen Chen, Central South University, Changsha, China, People's Republic of, Opher Baron, Ningyuan Chen

We investigate the daily collateral collecting problem of a central bank for clearing payment obligations among numerous banks who incur an opportunity cost of posting collateral and a borrowing cost from the central bank. The central bank chooses optimal collateral collection and debt offsetting policies in order to minimize the total expected costs of all banks. We formulate the problem as a two-stage stochastic program and derive a closed-form solution for the optimal collateral amount under standard linear cost assumptions. We extend the framework using cooperative game theory to analyze the benefits and stability of bank coalitions. Going beyond standard core allocation concepts from the cooperative game literature, we design a practical mechanism involving internal borrowing rules and a novel collateral allocation algorithm. We validate our theoretical findings and demonstrate the potential for significant system-wide cost savings using real daily interbank payment data from Payments Canada.

TB44

Building B Level 3 B304

Telehealth and multi-modal healthcare services

Invited Session

MSOM: Healthcare

Chair: Omer Berk Olmez, Baruch College, Graduate Center, CUNY, New York, NY, United States

Co-Chair: Alex Mills, Baruch College, New York, NY, 10010, United States

1 - Multi-Channel Healthcare Operations: the Impact of Video Visits on the Usage of in-Person Care

Tan (Suparek) Lekwijit, W. P. Carey School of Business, Arizona State University, Tempe, AZ, United States, Hummy Song, Christian Terwiesch, Krisda Chaiyachati

Healthcare organizations have increasingly adopted video visits as an alternative care channel that offers greater convenience and holds promise for improving patients' access to care. Despite the growing use, the impact of a new digital channel on care demand within existing physical channels has not been thoroughly investigated, especially in the primary care context of a traditional healthcare organization. We study a large healthcare system that made video visits available to a subset of its patients for primary care needs, prior to the COVID-19 pandemic. Using a difference-in-differences approach and visit data spanning over 9.5 million patient-months, we find that the introduction of video visits increases the demand for in-person primary care provider (PCP) visits by 20% and the demand for emergency department (ED) visits by 30%. When considering only patient-initiated visits, we find that the introduction of video visits increases the overall number of care initiations but does not affect in-person care initiations, suggesting that patients initiate care through video visits for care needs that may have otherwise gone unattended. Furthermore, we find that the increase in in-person demand arises primarily from patients who have poorer access to in-person care, as measured by distance and provider busyness. Lastly, compared to in-person visits, we find that video visits are less likely to provide care that can successfully conclude a care episode, especially when the medical condition is acute. Our findings have important capacity and revenue implications for healthcare organizations that are considering offering video visits as an additional care channel.

2 - The Impact of Information-Granularity and Prioritization on Patients' Care Modality Choices

Lin Zang, Stanford University, Stanford, CA, United States, Yue Hu, Ricky Roet-Green, Shujing Sun

The past few years have witnessed a significant expansion in telemedicine adoption by healthcare providers. On one hand, telemedicine has the potential to increase patients' access to medical appointments. On the other hand, due to the limitations of remote diagnostic and treatment methods, telemedicine may be insufficient for patients' treatment needs and may necessitate subsequent in-person follow-up visits. To better understand this tradeoff, we model the healthcare system as a queueing network providing two types of service: telemedicine and in-person consultations. We assume that an in-person visit guarantees successful treatment, whereas a telemedicine visit may fail to meet the patient's treatment needs with a probability that is contingent on individual patient characteristics. We formulate patients' strategic choices between these care modalities as a queueing game, and characterize the game-theoretic equilibrium and the socially optimal patients' choices. We further examine how improving patients' understanding of their telemedicine suitability through predictive analytics at the online triage stage affects system performance. We find that increasing information granularity maximizes the stability region of the system but may not always be optimal in reducing the average waiting time. This limitation, however, can be overcome by simultaneously deploying a priority rule that induces the social optimum under specific conditions. Finally, leveraging real-world data from a large academic hospital in the United States, we perform a comprehensive case study that encompasses both the development of a prediction model for in-person follow-up needs and the implementation of effective information provision and prioritization strategies.

3 - Balancing Health Outcomes and Patient Fatigue in Chronic Care

Wanting Yang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Jimmy Qin, Jingwei Zhang

We study the problem of scheduling appointments for patients with chronic diseases, incorporating the behavioral phenomenon of appointment fatigue, where a patient's no-show probability increases with their history of unnecessary false-positive recalls. We model this problem as a partially observable Markov decision process (POMDP) with the objective of maximizing finite-horizon health outcomes while mitigating the patient disengagement caused by fatigue. We analyze the structure of the optimal policy and establish several key properties. For the single-patient case, we prove the optimal policy is governed by a simple threshold on the provider's belief about the patient's health. We generalize this result to the N -patient setting, showing that the optimal policy possesses an elegant structure that partitions the high-dimensional state space into $N+1$ distinct regions, each corresponding to a unique optimal action. We develop three approximation policies: a theory-informed heuristic derived from our structural results, a Lagrangian relaxation-based policy, and an approximate linear programming (ALP) approach. Our work provides both deep structural insights and computationally tractable policies to help healthcare providers improve chronic care management by intelligently navigating the trade-off between clinical effectiveness and patient engagement.

4 - Understanding Customer Preferences for Virtual Versus in-Person Services

Alex Mills, Baruch College, New York, NY, United States, Omer Berk Olmez

As multi-channel services become more common, understanding the value customers place on virtual and in-person services has become increasingly important. Using data from a provider with in-person and virtual channels, we examine how service delivery channels and attributes—such as duration and distance—shape customer choices.

TB45

Building B Level 3 B305

Social Responsibility in Supply Chains

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Dwaipayan Roy, University of Virginia Darden School of Business, Charlottesville, VA, United States

1 - Towards An Operational Framework for Assessing Person-Job Fit

Bhavot Singh, Michigan State University, East Lansing, MI, United States, Sriram Narayanan, Rajiv Ranganathan, Hung Kuo

Inclusion of people with disabilities in the workforce is crucial for addressing labor shortages in the US. Despite comprising nearly 12% of the population, people with disabilities face significant barriers, however, current frameworks for fitting people with disabilities into jobs are inadequate. This study proposes a novel framework for inclusion.

2 - Spillover Benefits of Workplace Fairness on Manufacturing Operations: Evidence from Employment Non-Discrimination Acts in the United States

Max Ji, The Pennsylvania State University, University Park, PA, United States, Suvarat Dhanorkar, Kevin Linderman, Suresh Muthulingam

A growing body of work has examined how workplace policies promoting fair treatment positively influence workplace culture and the work environment. Beyond these foreseeable benefits, the impact of fair workplace initiatives on manufacturing processes remains underexplored. We seek to bridge this gap by leveraging a quasi-natural experiment that arises with the staggered adoption of Employment Non-Discrimination Acts (ENDAs), which prohibit discrimination based on characteristics such as gender, race or sexual orientation. Our results show that manufacturing facilities in states that have enacted ENDAs experience positive spillovers, as evidenced by significant reductions in their process waste. Our analysis reveals that this waste reduction effect operates through two separate pathways: operational pathway (i.e., based on the state where the manufacturing facility is located) and the management pathway (i.e., based on the state where the parent firm is headquartered). Our findings suggest that policies promoting workplace fairness can add value to manufacturing processes by fostering fair and merit-based practices.

3 - Labeling the Supply Chain: Signaling Supply Chain Transparency to Consumers Through Eco-Labels

Katelyn Thompson, University of South Carolina, Columbia, SC, United States, Blair Flicker, Priyank Arora

A central concern in operations and supply chain management is the processes by which products are made, an issue that has become increasingly salient among consumers. In response, many firms disclose their sustainability practices on product packaging through eco-labels, which are binary indicators of supply chain transparency. In the recent years, there is proliferation of not only the total number of labels in the market but also the number of labels on a product. Given the growing emphasis in practice on supply chain transparency, this paper explores how the use of multiple eco-labels affects consumers' perceptions of firm performance across different dimensions of sustainability, including environmental, social, and animal welfare. Specifically, we conduct incentivized studies to answer the following research question: For a firm that has already earned an eco-label pertaining to one dimension of sustainability, how does the addition of a second (or third) disclosure related to another dimension affect consumers' perception of the firm's performance on these multiple dimensions? Our work adds nuances to the understanding on supply chain transparency by examining interactions among multiple sustainability disclosures rather than evaluating each in isolation. By deepening the understanding of how multiple sustainability disclosures interact and how such interactions shape consumers' perceptions, this study offers practical insights for firms engaged in sustainable operations.

4 - Mobile Application with Real-TIME AI Exercise Feedback Improves Physiotherapy Adherence and Recovery

Pierre Gautreau, York University - Schulich School of Business, North York, ON, Canada, Phil Boyer, Moren Levesque, Helen Razmjou, Cristian Renteria, Ujash Sheth, Robin Richards, Jay Shaw, Cari Whyne, David Burns

This study investigates the impact of the Smart Physiotherapy Activity Recognition System (SPARS) and its companion mobile application, PT+ on patient exercise adherence and recovery outcomes in shoulder physiotherapy. SPARS uses inertial sensors in commercially available smartwatches to track the quantity and quality of exercise performed. PT+ provides real-time feedback on exercise technique and monitors patient progress. Our comparative cohort study followed participants for 12 weeks. The intervention group (N=24) received a smartwatch and tablet equipped with SPARS/PT+, while the control group (N=100) received a smartwatch that passively tracked physical therapy participation using SPARS without displaying results. 6,498 patient-days of exercise data and 2,093 hours of movement records indicates that the intervention group performed significantly more exercise and had better recovery outcomes. Analysis of technique adherence and patient interviews revealed that sharing the SPARS/PT+ feedback with a physiotherapist was a key motivator, leading to increased exercise participation and improved recovery.

5 - Pack Size and Pricing in a Developing Economy: the Strategic Role of Small Packs

Anant Mishra, Carlson School of Management, University of Minnesota, Minneapolis, MN, United States, Arzi Adbi, Ajay Bhaskarabhatla, Jingwen Yang

Although small-packs, originally designed to serve the base of the pyramid (BOP), are ubiquitous, we know very little about how manufacturers price these offerings relative to their large packs. Conceptual works drawing from quantity discount models predict a per-unit price premium on small packs, reflecting the economies of scale in production and distribution. Yet liquidity constraints and limited consumer awareness associated with the BOP challenge this logic. Our study explores the interplay between pack size and pricing by analyzing how manufacturers price their small-packs relative to large-packs.

TB46

Building B Level 3 B306

Robust Dynamic Decision Making

Invited Session

Decision Analysis Society

Chair: Zhengchao Wang, Imperial College London, London, United Kingdom

Co-Chair: Fupeng Sun, Imperial College London, 810 The Robinson West, 1 Atlantic Crescent, London, HA9 0TT

1 - Dynamic Adherence Markov Decision Process

Zhengchao Wang, Imperial College London, London, United Kingdom, Huikang Liu, Huan Xu

Algorithmic recommendations are seldom followed to the letter: real-world users may only partially adopt, delay, or abandon suggested actions as their trust fluctuates. Adherence-aware Markov decision processes (AAMPDs) address this gap by modelling the implemented policy as a stochastic blend of the algorithmic recommendation and a baseline strategy. We extend this framework to settings in which adherence evolves over time and introduce Dynamic Adherence-Aware MDPs (DAAMPDs), where a latent trust variable governs the probability of compliance. Our first contribution shows that a DAAMPD can be reformulated as a conventional MDP with augmented state, yielding transition and reward kernels that embed both trust dynamics and user inertia. Second, we develop efficient algorithms for computing optimal recommendation policies within the extended state space. Finally, we prove that, under a mild regularity condition on outcome-driven trust updates, the decision maker's adherence level converges almost surely to one, guaranteeing eventual full adoption of the learned policy. The proposed framework bridges the gap between prescriptive analytics and human behaviour, providing a principled basis for algorithms that must earn and retain user trust over time.

2 - Risk-Averse Total-Reward MDPs with Erm and Evar

Marek Petrik, University of New Hampshire, Durham, NH, United States, Xihong Su, Julien Grand-Clement

Optimizing risk-averse objectives in discounted MDPs is challenging because most models do not admit direct dynamic programming equations and require complex history-dependent policies. In this paper, we show that the risk-averse $\{\text{total reward criterion}\}$, under the Entropic Risk Measure (ERM) and Entropic Value at Risk (EVaR) risk measures, can be optimized by a stationary policy, making it simple to analyze, interpret, and deploy. We propose exponential value iteration, policy iteration, and linear programming to compute optimal policies. Compared with prior work, our results only require the relatively mild condition of transient MDPs and allow for $\{\text{both}\}$ positive and negative rewards. Our results indicate that the total reward criterion may be preferable to the discounted criterion in a broad range of risk-averse reinforcement learning domains.

3 - Strongly-Polynomial TIME and Validation Analysis of Policy Gradient Methods

Caleb Ju, Georgia Institute of Technology, Atlanta, GA, United States

We propose a novel termination criterion, termed the advantage gap function, for finite state and action Markov decision processes (MDP) and reinforcement learning (RL). By incorporating this advantage gap function into the design of step size rules and deriving a new linear rate of convergence that is independent of the stationary state distribution of the optimal policy, we demonstrate that policy gradient methods can solve MDPs in strongly-polynomial time. To the best of our knowledge, this is the first time that such strong convergence properties have been established for policy gradient methods. Moreover, in the stochastic setting, where only stochastic estimates of policy gradients are available, we show that the advantage gap function provides close approximations of the optimality gap for each individual state and exhibits a sublinear rate of convergence at every state. The advantage gap function can be easily estimated in the stochastic case, and when coupled with easily computable upper bounds on policy values, they provide a convenient way to validate the solutions generated by policy gradient methods. Therefore, our developments offer a principled and computable measure of optimality for RL, whereas current practice tends to rely on algorithm-to-algorithm or baselines comparisons with no certificate of optimality.

4 - Robust Weakly Coupled Markov Decision Processes

Fupeng Sun, Imperial College Business School, London, United Kingdom, Huikang Liu, Zhe Liu, Wolfram Wiesemann

We study Robust Weakly Coupled Markov Decision Processes (RWCMDPs), a modeling paradigm for sequential decision-making under uncertainty in the transition kernel, where the global problem decomposes into independent subproblems coupled through linking constraints. The objective is to maximize the expected reward under the worst-case realization of the uncertainty for the finite horizon. We develop a general solution framework based on robust Lagrangian relaxation with state-action ambiguity sets. We prove that the resulting robust Lagrangian dual problem is non-convex and admits three equivalent formulations. These formulations enable the derivation of upper and lower bounds via subgradient methods and constant decision rule approximations. In a data-driven setting—where the ambiguity set is centered at the maximum likelihood estimator and converges in infinity-norm to the true transition kernel—we prove that the resulting policies constructed from both the upper and lower bounds are asymptotically optimal for the weakly coupled MDP under the true model, as both the ambiguity set radius and the number of subproblems grow large. Numerical experiments demonstrate the superior performance of our proposed methods.

5 - Fast Convergent Risk-Aware Reinforcement Learning

Junjie Lei, The University of Hong Kong, Hong Kong, China, People's Republic of, Wenjie Huang, Dionysios Kalogerias

This work proposes a fast-converging reinforcement learning algorithm under risk-aware objectives, termed Zap Risk-aware Q-learning (Zap RaQL). By integrating elicitable and coherent risk measures—specifically, expectiles—into the Q-learning framework, we effectively capture the downside risk in Markov decision processes (MDPs). The algorithm employs a two-timescale stochastic approximation scheme combining Zap Q-learning with a Zap-style Newton-Raphson method for risk estimation, both of which benefit from second-order updates. We prove that the proposed algorithm achieves almost sure convergence. Theoretical analysis is supported by numerical experiments demonstrating that Zap RaQL significantly reduces the variance of Bellman errors compared to standard risk-aware Q-learning. Our work provides a principled and efficient framework for risk-sensitive decision-making in reinforcement learning.

TB47

Building B Level 3 B307

Assortment Optimization and Retail Strategy

Contributed Session

Contributed

Chair: Keji Wei, Tongji University, Shanghai, N/A, China, People's Republic of

1 - Assortment Optimization with Customer-Centric Fairness

Yihua He, The Hong Kong University of Science and Technology, Hong Kong, China, People's Republic of, Qian Liu, Jin Qi

The rapid growth of e-commerce platforms has enabled businesses to implement customized assortment strategies, tailoring product offerings to different customer groups. While these strategies can significantly enhance profitability, they often raise concerns about fairness, particularly when certain customer groups receive fewer options. Such practices can provoke customer backlash and damage a company's reputation. Motivated by these issues, we propose a fairness-aware framework for assortment optimization that balances revenue maximization with equitable treatment of customer groups. By introducing cardinality-based fairness constraints, our framework ensures that sellers provide a similar number of products to different groups, subject to varying levels of fairness tolerance. Our analysis demonstrates that fairness constraints compel sellers to offer more products to disadvantaged groups while reducing the advantages previously enjoyed by privileged groups. In the meantime, the fairness constraints also enhance the discriminated group's customer surplus even though the seller does not explicitly target this outcome. On the other hand, the sellers have no incentive to exceed the minimum fairness requirements. Moreover, we develop efficient polynomial algorithms to solve the problem exactly when customers are categorized into two categories and for the multi-type scenarios with semi-independent tolerance levels. We also propose an approximate algorithm for the most general multi-type scenarios with a guaranteed performance bound. Numerical results show that fairness constraints have a limited impact on seller revenue, making them less likely to resist such regulations. However, policymakers must carefully weigh the trade-offs as fairness constraints can reduce overall customer surplus in some cases.

2 - Optimizing Tiered Product Pricing with Choice-Based Demand and Capacity Constraints

Sai Krishna Reddi, University at Buffalo, Buffalo, NY, United States

We present an optimization framework for pricing and designing tiered products using a discrete-choice demand model. The approach maximizes expected profit by jointly optimizing product prices and feature configurations, while accounting for visit frequency, feature costs, and ancillary revenues. The model incorporates capacity constraints and feature hierarchy, supporting interpretable, profitable designs across subscription or access-based industries.

3 - Bad Store or Bad Inventory? a Structural Model of Multi-Location Retail Performance with Assortment Variation

Katherine Ashley, Northeastern University, Boston, MA, United States

Fashion retailers with multiple brick-and-mortar store locations face two challenges simultaneously: they must select an appealing range of products to make or buy from garment manufacturers, and they must allocate available units of inventory across store outlets to match supply with uncertain demand in multiple markets. Yet once these decisions have been made, a further challenge remains: fairly evaluating the outcomes that are observed in different stores that received inventory allocations of varying consumer appeal. While extensive research examines assortment optimization and inventory allocation decisions, the existing literature does not focus on disentangling store-specific operational efficiency from product-specific appeal in a large-scale empirical setting.

Using data from a U.S. fashion retailer, we propose a simultaneous equations model that allows for the estimation of store effects and product effects on item-level profitability. Our identification strategy exploits cross-sectional variation in product availability: items sold in multiple locations enable separation of inherent product appeal from store-specific factors, while stores carrying overlapping assortments allow identification of location-specific operational capabilities. The framework accounts for potential endogeneity in allocation decisions, where better-performing stores may receive more appealing inventory.

This research extends structural estimation techniques in operations management and provides an econometric framework designed specifically for the multi-location retail attribution problem. The results of this model can inform store- and chain-level assortment planning, inventory allocation, and store closing decisions with separate insights into product performance and store-level execution.

4 - Assortment Planning and Free Shipping Strategies Under the Multinomial Logit Choice Model

Sahika Sahan Konur, Texas State University, San Marcos, TX, United States, Hafila Morais Max Morais

In this study, we examine the integrated problem of assortment planning, free shipping strategies, and shipping fee settings under the Multinomial Logit (MNL) model. We first establish that the general version of the problem is NP-hard. To explore more practical approaches, we analyze the problem across nine scenarios formed by the combination of three free shipping strategies—all-or-none, threshold-based, and selective free shipping—and three shipping fee policies: box-rate, flat-rate, and discounted box-rate. For most of these scenarios, we are able to characterize the optimal solution. However, when selective free shipping is combined with either flat-rate or discounted box-rate fees, the problem remains NP-hard. For these two challenging cases, we propose pseudo-polynomial ϵ -approximation algorithms. Finally, we conduct a series of numerical experiments to investigate the conditions under which each scenario performs best, as well as to assess the impact of retailers' shipping costs and customers' sensitivity to shipping fees on overall revenue.

5 - Product Line Design Under a Context-Dependent Multinomial Logit Model

Keji Wei, Tongji University, Shanghai, China, People's Republic of, Xiaoyun Niu, Yufeng Cao

When an airline passenger selects among flights departing at similar times, the availability of nearby alternatives can subtly alter the perceived appeal of each option—a phenomenon that traditional discrete choice models often overlook. Using a proprietary dataset of airline sales, we empirically show that passenger utility is shaped not only by flight attributes but also by the density of competing options and their temporal proximity.

Building on these insights, we develop a **product line design framework** under a **context-dependent multinomial logit model**, where the attractiveness of each flight dynamically adjusts based on the offered set. We apply this framework to **airline fleet assignment and timetabling decisions**, explicitly modeling how product availability and scheduling influence customer choice and generate cannibalization effects between competing flights. To address the computational complexity introduced by context-sensitive interactions, we propose **mixed-integer reformulations** and **tractable solution techniques**.

Our work advances both **empirical understanding of context effects** in choice behavior and **methodological innovation in revenue management optimization**, enabling more sophisticated decision-making under flexible, context-dependent utility models.

TB48

Building B Level 3 B308

Advances in Decision Analysis

Invited Session

Decision Analysis Society

Chair: Manel Baucells, Darden School of Business, Charlottesville, VA, 22903, United States

1 - On the Properties of the Metalog Distribution

Manel Baucells, Darden School of Business, Charlottesville, VA, United States

There has been a longstanding interest among statisticians and analysts for parametric probability distributions flexible enough to match arbitrarily shaped, non-normal data sets (e.g., the Pearson family, or the Johnson family). The metalog quantile function emerged in 2016 as a new alternative and has been widely adopted for use in practice for its simplicity and flexibility. Important properties of the metalog such as moments and modes or how to address infeasibility, remain poorly understood. We introduce the redefined metalog 2.0 which provides guarantees of non-degeneracy. We establish the possible number of modes, and uncover elementary expressions for the moments and the partial expectations. We also show a new way to control the support of the distribution, a precise diagnosis of feasibility and—in case of infeasibility—an efficient algorithm to estimate the best feasible fit. We conclude by showcasing some applications.

2 - A Walk Down Memory Lane: How Memories Influence Stock Investment and Information Processing

alessandra cillo, LIUC University, Castellanza (Va), Italy, Giovanni Burro, Sergio Alessandro Castagnetti, Giovanni Crespi

We study the impact of memories on investment decisions and how they affect the way individuals use stock-related information. In the first experiment, we find that eliciting memories before an investment task polarizes the level of the amount invested: positive memories increase the level invested compared to recalling non-positive ones. In the second experiment, we elicit memories of the investment from all participants before the investment, but only some of them receive positive and truthful financial information about the stock they are about to invest in. While information provision increases investment, it only does so for those who do not recall personal memories, as opposed to personal ones. We rationalize our findings in light of a memory-based model.

3 - Truthful Budget Aggregation: Beyond Moving-Phantom Mechanisms

Rupert Freeman, University of Virginia, Charlottesville, VA, United States, Mark de Berg, Ulrike Schmidt-Kraepelin, Markus Utke

We study a budget-aggregation setting in which a number of voters report their ideal distribution of a budget over a set of projects, and a mechanism aggregates these reports into an allocation. In the first part of this work, we study the space of truthful mechanisms, i.e., those mechanisms that do not incentivize voters to misreport their preferences. For the case of two projects, the set of mechanisms that are truthful and additionally meet a range of basic desiderata (anonymity, neutrality, and continuity) exactly coincides with the so-called moving-phantom mechanisms, but whether this space is richer for more projects was repeatedly stated as an open question. We answer this question in the affirmative by presenting a class of truthful mechanisms that are not moving-phantoms. In the second part, we explore the interplay between truthfulness and fairness. On the negative side, we show that there exist instances where all truthful mechanisms must disregard the preferences of half the electorate, violating key fairness principles. On the positive side, we propose a novel moving-phantom mechanism and show that it provides project fairness guarantees that are optimal among all truthful mechanisms. Project fairness uses the mean allocation as a baseline and quantifies the maximum deviation of funding received by any project from this baseline. As a corollary of our results, we show that our new mechanism minimizes the worst-case ℓ_1 distance to the mean for three projects and gives the first non-trivial bounds on this quantity for more than three projects.

4 - Optimal Index Policies for Sequential Search: a Unified Framework Under Bayesian Learning

Zixin (Stephen) Xu, Darden Business School, University of Virginia, Charlottesville, VA, United States, Manel Baucells, Sasa Zorc

Sequential search models are critical tools in operations management, with extensive applications ranging from job hunting and consumer purchasing to technology adoption. However, traditional models often unrealistically assume complete knowledge of underlying distributions. Addressing this limitation, we develop a unified Bayesian framework for optimal sequential search with unknown distribution parameters, applicable to diverse recall structures, risk preferences, and discounting behaviors. Our approach introduces a general family of distributions, classified into subfamilies based on parameter types—location, scale, or both—and derives an optimal index policy valid across these subfamilies. This policy simplifies complex dynamic programming problems by using standardized sufficient statistics, thus maintaining tractability even when incorporating different utility transformations (CARA, CRRA) and discounting methods (exponential, hyperbolic). We further disprove a longstanding conjecture by showing that, contrary to prior belief, search without recall under exponential families and conjugate priors need not follow a single-reservation-price rule. Importantly, our framework offers immediate applications in the computer science community, particularly in sequential sampling problems such as Bayesian optimal stopping for Large Language Model (LLM) inference. Beyond its theoretical contributions, our method enables robust and adaptive sampling strategies under uncertainty, bridging behavioral modeling and machine learning. This result significantly expands the theoretical and practical understanding of sequential search, providing a versatile, computationally efficient methodology that captures a wide range of behavioral, economic, and computational scenarios.

TB49

Building B Level 3 B309

Innovating for Impact: Technology, Sustainability, and Social Good

Invited Session

MSOM: Technology, Innovation, and Entrepreneurship

Chair: Yuan Shi, MIT, Boston, 02110, United States

Co-Chair: Yanchong Zheng, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Optimizing Electric Vehicle Charging Infrastructure on University Campuses: a Queueing-Theoretic Analysis of Capacity and Congestion**Yikai Zhang, Boston University, Boston, MA, United States, Jinglong Zhao, Justin Ren**

As the transition to electric vehicles (EVs) accelerates in both urban and academic environments, effective management of EV charging infrastructure is becoming increasingly critical. This study evaluates the operational performance of a small-scale EV charging station equipped with limited chargers, addressing issues of capacity constraints and user congestion through a queueing-theoretic framework and Monte Carlo simulations. Empirical data reveal high blocking probabilities that prevent users from accessing charging services, highlighting substantial user inconvenience. Across multiple policy scenarios varying in charger availability and pricing strategies, the analysis finds the policy which minimizes total annual cost and alleviates current congestion. This scalable framework provides actionable insights for decision-makers managing EV infrastructure in campuses or other small-scale settings, contributing to broader sustainable transportation efforts.

2 - Encoding and Decoding Illegal Wildlife Trade Networks with Machine Learning and Centrality Analysis**Hannah Murray, University of Southern California, Los Angeles, CA, United States, Meredith Gore, Bistra Dilkina**

The illegal wildlife trade is a key driver of biodiversity loss and a barrier to desired transformations in socio-environmental systems. It is known to exploit licit networks, such as the global airline flight network, yet the ability of science to support efforts to reduce the illegal wildlife trade remains underdeveloped. Research on the illegal wildlife trade relies on biased aggregate counts of observed incidents, leading to potential policy misguidance. Here we utilize centrality analysis and predictive modeling to encode and decode illegal wildlife trade flight networks to improve sensemaking with implications for sustainable futures. Methods advance existing analyses by revealing traffickers prioritize airports with high centrality in the global airline flight network, high incidence of flora crimes, and a limited level of non-state actor oversight. Machine learning models identify airports likely to be involved in the illegal wildlife trade not currently implicated in seizure data, including in the United States.

3 - Evidence of Spillovers from (Non)Cooperative Human-Bot to Human-Human Interactions**Margaret Traeger, University of Notre Dame, Notre Dame, IN, United States, Ashley Harrell**

It is well-documented that cooperation spills over among humans: people's cooperative choices are influenced by their (non)cooperative alters, even in downstream interactions with new partners. We ask: do (non)cooperative interactions with bots spill over to subsequent interactions with humans—and if so, how? Across two pre-registered experiments (combined N = 83,411 decisions by 4,171 participants told they are interacting with a bot or human), we demonstrate that human-bot interactions do spill over to human-human ones, and in two ways. First, interacting with a bot reduces cooperation not only during the initial interaction; it also reduces downstream cooperation toward a new human partner. Additionally, bots' (like human partners') behavior matters: interactions with bots playing tit-for-tat promote cooperation, and interactions with noncooperative bots reduce cooperation, toward downstream human partners. The implementation of bots in previously human-only spaces alters human cooperation, not just toward bots, but toward other humans as well.

4 - Tip Your Farmer? Implications of Tipping on Smallholder Welfare in Agricultural Supply Chains**Basak Kalkanci, Georgia Institute of Technology, Atlanta, GA, United States, Saeed Alizamir**

An emerging financial innovation enabled by technological advancements in agricultural supply chains is the ability to “tip the farmers”. This mechanism empowers socially-conscious consumers to identify and support the individual farmers of their products by sending the direct payments, or tips. We shed light on the implications of tipping through a game-theoretic model that captures the interactions between a mass of infinitesimal farmers, a population of socially-conscious consumers, and an agricultural firm who plays the intermediary role between farmers and consumers. We find that if tipping is implemented under the right conditions, it can create a triple-win for all supply chain members including every individual farmer. This happens when farmers' outside option takes moderate values and the consumers' social-consciousness is sufficiently high. When these conditions are violated, however, farmers and/or consumers could be worse off in the presence of tipping. Furthermore, even in situations where farmers benefit from tipping in expectation, it may entail disparity and exacerbates inequity in the farmer population, which is undesirable from a social responsibility standpoint. Adverse outcomes such as these occur in parameter regions wherein the firm exploits the consumers' tip to lower its wholesale price commitment to the farmers. Thus, caution must be exercised in implementing a tipping program as it may lead to reduction in farmers' expected and/or realized income as well as consumer welfare. Even when the firm commits to maintaining the same wholesale price a priori, tipping can be detrimental to the firm while helping consumers and farmers.

TB50

Building B Level 3 B310

Operational Advances in Immigration

Invited Session

MSOM: Service Operations

Chair: Daniel Freund, MIT, Cambridge, MA, United States

Co-Chair: Andrew Trapp, Worcester Polytechnic Institute, Worcester, MA, United States

1 - Dynamic Matching with Post-Allocation Service and Its Application to Refugee Resettlement**Vahideh Manshadi, Yale, New Haven, CT, United States**

Motivated by our collaboration with a major refugee resettlement agency in the U.S., we study a dynamic matching problem where each arrival (a refugee case) must be matched immediately and irrevocably to one of the static resources (a location with a fixed annual quota). In addition to consuming the static resource, each case requires post-allocation service from a server, such as a translator. Given the time-consuming nature of service, a server may not be available at a given time, thus referred to as a dynamic resource. Upon matching, the case waits to avail service in a first-come-first-serve manner. Bursty matching to a location may result in undesirable congestion at its corresponding server. Consequently, the planner (the agency) faces a dynamic matching problem with an objective that combines matching reward (pair-specific employment outcomes) with costs for congestion at dynamic resources and over-allocation of static ones. Motivated by observed fluctuations in refugee composition across years, we design algorithms without relying on distributional knowledge from past years' data. We develop learning-based algorithms that are asymptotically optimal in certain regimes, interpretable, and computationally fast. Our design is based on learning the dual variables of the underlying optimization problem; however, the main challenge lies in the time-varying nature of dual variables associated with dynamic resources. To address this, our theoretical development draws on Lyapunov analysis, adversarial learning, and stochastic optimization. When tested on real data from the partner agency, our method outperforms existing ones, making it a viable candidate for replacing current practices upon experimentation.

2 - Regulating Wait-Driven Requests in Queues**Daniel Freund, MIT, Cambridge, MA, United States**

Rational queueing has fundamentally relied on the assumption that service requests lose utility while waiting in queues. This paper is motivated by applications where this assumption fails to hold and certain requests actually benefit from waiting. One such application is the US affirmative asylum system, for which it has been suggested that long wait times incentivize non-meritorious applications. We study the impact of scheduling policies in such a setting through a dynamical system where, in each round, a given scheduling policy and a realized request rate determine the wait time distribution in a fluid queueing system. Observing the expected benefit from waiting in one round, requests update their decisions for the next round. Assuming (i) a concave benefit function from waiting and (ii) that the distribution of abandonment times has an increasing hazard rate, we prove that Last-In-First-Out (LIFO) minimizes while First-In-First-Out (FIFO) maximizes the backlog among all work-conserving policies. Moreover, we show that the dynamical system exhibits metastability: for either FIFO or LIFO, the system converges, through two different types of herd behavior, to either a zero-wait or a congested equilibrium. In practice, LIFO is often critiqued for its failure to maintain fairness: earlier requests should get earlier service. Our results demonstrate a fairness-efficiency trade-off between LIFO and FIFO, but also illustrate limitations of hybrid policies, which probabilistically follow either LIFO or FIFO to navigate this trade-off. Upon extending a known fairness notion to queueing systems with abandonment, we demonstrate that LIFO may Pareto-dominate hybrid policies with respect to

3 - Alignment and Misalignment of Fairness Constraints in Candidate-Job Matching

Marcela Vasconcellos, Worcester Polytechnic Institute, Worcester, MA, United States, Swati Gupta, Andrew Trapp

As AI-driven systems increasingly mediate labor market access, concerns about fairness and bias in candidate to job matching have intensified. We examine how fairness constraints align with the true compatibility signals in predictive candidate-to-job matching systems. Specifically, we investigate whether fairness interventions intended to equalize opportunities across groups move predicted matching scores closer to true candidate to job compatibility, or instead introduce further misalignment. Starting from a machine learning model that estimates candidate-to-job fit, we account for bias in these predictions arising from taste-based or statistical discrimination. We examine the effect of adjusting fit scores using weights from dualized fairness constraints of a bipartite matching formulation solved via linear programming. We show that, in some cases, these constraints can correct for taste-based and statistical discrimination. We evaluate our approach on real-world data from a social startup focused on improving equitable employment access, measuring how closely the interventions align with real-world weights via group-specific accuracy, precision, and error rates. Our findings identify when fairness constraints are aligned or misaligned with actual compatibility signals, providing practical guidance for designing equitable AI systems in hiring.

4 - Data Envelopment Analysis Meets Private Refugee Resettlement

Fatemeh Farajzadeh, Worcester Polytechnic Institute, Worcester, MA, United States, Sarah Stanlick, Vincent Charles, Andrew Trapp

This study introduces a data-driven framework to evaluate success factors in private refugee resettlement, with a focus on Neighborhood Support Teams (NSTs). These community-based groups supported refugees admitted under the humanitarian parole program for private resettlement in the United States from late 2021 through 2022. We develop a customized Data Envelopment Analysis (DEA) to assess NST performance, informed by targeted data collection capturing key influencing factors and the complexity of their operations. To address the sensitivity of DEA to variable selection, we develop an iterative method combining logistic regression and statistical significance testing. The results provide actionable insights for policymakers and resettlement agencies to enhance decentralized refugee support through evidence-based decision-making.

TB51

Building B Level 3 B311

Empirical Retail Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Xabier Barriola, Tilburg University, Tilburg, Netherlands

1 - Incentive Design for Long-Term Retention of New Gig Workers

Natalie Epstein, Harvard Business School, Cambridge, MA, United States, Ignacio Rios

High worker turnover remains a persistent challenge for platform-based businesses. Not only does it increase acquisition and training costs, but it also negatively impacts service quality. In this paper, we explore how the design of short-term monetary incentives impacts worker retention and engagement in the long term.

We gather data from a B2B platform that offers last-mile delivery services to a beverage conglomerate in Brazil. High worker turnover and different threshold-based incentive structures across markets make this an ideal setting. We analyze the company's change in incentive design using a difference-in-differences (DiD) model and show that when the incentive compensates workers on the number of deliveries in the extra route (intensive margin) as opposed to just on taking an extra route (extensive margin) they are more likely to work until they reach the bonus and to continue working more intensively even after the bonus period. We show that this long-lasting effect is not entirely driven by earning more money through the bonus, and discuss potential mechanisms that can drive this behavior, such as early learning, inertia, and goal setting.

Monetary incentives are a valuable tool to increase payment in return for effort and engagement. The design of such incentives is not straightforward, and the structure of short-term incentives can increase worker engagement and have lasting effects in the long run. Providing incentives early on and structuring them to induce such behavior can turn into a valuable and low-cost lever for platforms that seek to increase worker retention.

2 - Algorithms in Retail Marketplaces

Nil Karacaoglu, The Ohio State University, Columbus, OH, United States, Santiago Gallino, Antonio Moreno

This paper examines how algorithmic recommendations influence customer behavior in online marketplaces using proprietary data from a leading digital marketplace.

3 - Digital Platforms and the Fissured Workforce - Consequences of Temporary Workforce Strategies in Service Operations

Jose Lopez, Texas A&M University, College Station, TX, United States

Service organizations frequently rely on temporary workers to manage fluctuating workloads, leveraging short-term flexibility and reduced fixed costs. However, this strategy may inadvertently lead to long-term operational inefficiencies. Our study examines the impact of temporary labor strategies on service performance, workforce productivity, and operational stability using a dynamic model in a logistics-oriented service context. We study four main sectors: processing and shipment, workforce dynamics (including rookies, experienced temporary, and permanent employees), capacity adjustments based on workload pressures, and performance feedback mechanisms. Simulations informed by empirical data from a cross-docking logistics provider highlight critical trade-offs. Heavy reliance on temporary labor results in continuous inefficiencies due to high turnover and perpetual training needs, negatively affecting productivity. Mixed labor approaches offer moderate flexibility but still incur elevated management overhead and decreased workforce cohesion. Findings underscore managerial tendencies to underestimate turnover's compounded productivity impacts.

4 - Spillover Effects of Labor Disruptions

Xabier Barriola, Tilburg University, Tilburg, Netherlands

Grocery stores provide an essential and highly competitive service in many communities. Hourly employees play a critical role in the provision of this service through a wide range of functions, including receiving, stocking, checkout, and customer care. We examine the effects on the customer experience, store performance, and supplier performance when these functions are disrupted through an organized strike.

TB52

Building B Level 3 B312

Business Model Innovations for Sustainable Operations

Invited Session

MSOM: Sustainable Operations

Chair: Luyi Gui, The Paul Merage School of Business, UC Irvine, Irvine, CA, United States

Co-Chair: Xi Henry Lin, UC Irvine, Paul Merage School of Business, Irvine, United States

1 - Resale-as-a-Service: The Impact of Second-Hand Platforms on Fashion Goods Retailer**Yuan Guo, George Washington University, Washington, DC, United States, Fernando Bernstein, Robert Swinney**

Fashion resale has evolved from peer-to-peer exchanges to platform-mediated services, prompting brands to reassess the role of secondhand markets. We study a setting where resale is offered as a service (RaaS) via a platform that intermediates used product flows. Our model captures the interaction between a fashion retailer and the resale platform under a passive resale design, where returned goods are accepted and resold by default.

We analyze how resale access reshapes the retailer's assortment and pricing strategies in the primary market, and identify conditions under which resale acts as a liability, neutral byproduct, or strategic complement. The model reveals how resale margins, return composition, and consumer taste shifts jointly determine the retailer's optimal design. We also explore how platform coordination through revenue-sharing affects market outcomes and incentives.

Our findings offer insights into the operational design of RaaS partnerships, and suggest directions for future work on platform-driven control in fashion recommerce.

2 - Dynamic Battery Swapping and Charging in Large-Scale Electric Vehicle Networks**Zuguang Gao, UC Irvine, Irvine, CA, United States, Deniz Akturk**

This talk addresses the daily operational challenges faced by electric vehicle (EV) companies managing large-scale battery swapping stations in urban environments, where meeting customer demand while optimizing charging strategies is a critical objective. We present a novel stochastic programming model that accommodates multiple charging levels and accounts for uncertain demand across discrete time periods. By reformulating the underlying Markov Decision Process (MDP), we propose a mean-field approximation that leverages aggregated system behavior to reduce computational complexity. The key insight of our approach is that, as the number of swapping stations grows, the mean-field policy remains asymptotically near-optimal, with a sublinear deviation from the true optimal cost. To enforce feasibility after rounding decisions, we develop a systematic procedure that ensures the resulting solution respects capacity constraints and battery inventory limitations. This research contributes to the growing body of literature on decision-making under uncertainty in battery management systems, offering valuable insights for both academic researchers and industry practitioners seeking scalable and efficient operational strategies in emerging urban mobility ecosystems. We also incorporate theoretical guarantees to show that the total cost incurred by the proposed policy grows sublinearly in the number of stations. This ensures scalable performance for large EV networks under diverse operational conditions.

3 - Reduce and Reveal: Emissions Abatement and Disclosure in Supply Chains**Vishal Agrawal, Georgetown University, Washington, DC, United States, Soudipta Chakraborty**

To reduce their exposure to climate change risks, investors increasingly factor firms' emissions into their investment decisions. Consequently, firms exert effort to abate emissions and disclose these emissions annually. When disclosing emissions is voluntary, firms are often strategic in their reporting: some engage in greenwashing—portraying themselves as more sustainable than they truly are—while others participate in greenhushing—appearing less sustainable than they actually are. Such strategic behavior has led to growing calls for regulations mandating accurate emissions disclosure. We study firms' emissions abatement and reporting strategies under both voluntary and mandatory disclosure. Our findings reveal that the disclosure strategy significantly influences firms' abatement efforts. Contrary to common intuition, we demonstrate that although mandatory disclosure ensures accurate emissions reporting, firms sometimes exert greater abatement effort under voluntary disclosure, even when they engage in greenwashing.

4 - Flock Management for Solar Grazing**Vincent Slaugh, Cornell University, Ithaca, NY, United States, Linwei Xin, Shane Henderson**

To control vegetation, solar asset owners employ farmers to graze sheep on solar sites. While earning revenue from vegetative maintenance, farmers have the opportunity to provide agricultural value through raising and selling sheep. Accounting for flock dynamics of births, deaths, and sales, we model the operational economics of solar grazing for farmers and focus on key flock management decisions.

We develop analytical models that determine how many sheep farmers should buy initially and how to manage flock ages over the long term. First, we show that a farmer's total profit—including revenue from mowing contracts and livestock sales—behaves as a piecewise linear and concave function of the number of sheep purchased. Second, we introduce an age-based model showing that a simple "maximum-lifetime" approach—i.e., keeping sheep for their entire productive lifetime before culling matches the performance of more complicated strategies. We also find that as a flock grows large, uncertainty (e.g., birthrate variations) has minimal impact if farmers retain young ewes consistently and cull only in their final productive year.

Our results help small-scale farmers, renewable energy developers, and policymakers see how solar grazing enhances the environmental benefits of solar farms and preserves agricultural value in land converted to solar farms. Through a partnership with a solar grazing platform that helps solar grazers and connects them with site owners, a spreadsheet implementation of our model has helped dozens of farmers make procurement decisions, predict cash flows, and even procure financing to purchase a flock.

5 - The Environmental and Social Implications of Pay-What-You-Want Food Waste Supermarkets**Xi Henry Lin, UC Irvine, Paul Merage School of Business, Irvine, CA, United States, Luyi Gui, Luyi Yang**

Around 1.3 billion tons of food are wasted globally each year, accounting for about one-third of all food produced for human consumption; nevertheless, a substantial portion of the global population faces moderate or severe food insecurity, indicating a significant disparity in food distribution.

Food donation, particularly at the retail level, has become an important approach to recovering surplus but edible food. In addition to traditional donation channels, food waste supermarkets—retail outlets selling donated food at low prices—have gained traction. These outlets adopt either posted pricing or the more flexible "pay-what-you-want" (PWYW) model, the latter of which is often advocated for its social and environmental benefits. However, it remains unclear whether PWYW can consistently deliver such benefits.

This study develops a game-theoretic model to analyze the interplay between a regular retailer's procurement decision, consumer purchase behavior, and the impact of different pricing schemes. We show that the difference in procurement quantities between the two pricing schemes is driven by consumer characteristics, such as the size of the food-insecure population and their sensitivity to advantageous inequality. Under PWYW, moderate sensitivity may lead to higher procurement and increased food waste, while low or high sensitivity may result in lower procurement and reduced food access.

Overall, our study indicates that the common belief that PWYW can reduce food waste and improve access for the food-insecure community simultaneously does not always hold: depending on consumer characteristics, there are cases where the environmental benefit comes at the expense of the social benefit, and vice versa.

TB53

Building B Level 3 B313

Recent Advancements in the Practice of Revenue Management and Pricing

Invited Session

Revenue Management

Chair: Ezgi Eren, PROS, Inc., Houston, 77098, United States

1 - Real-Time Competitor-Aware Dynamic Pricing for Airline Revenue Management

Yuxuan Lu, Sabre, Boston, MA, United States

Customer purchase decisions hinge not only on an airline's own fares but also on the competitor prices shown at shopping time. Yet many carriers still rely on judgment-based manual overrides to react to those competitive moves. We introduce an end-to-end framework that embeds real-time competitive intelligence directly into dynamic pricing.

The approach has two pillars. First, we estimate a multinomial-logit passenger choice model that treats every itinerary—host and competitor alike—as an option in the customer's consideration set, explicitly including a “do-not-fly” outcome when the offered prices exceed the traveler's budget. Second, we embed this choice model in an optimizer that recalculates the host airline's fare in real time, maximizing expected revenue while fully accounting for competing itineraries and their prices.

2 - A Win-Only Logistic Model for B2B Pricing Decisions

Zijian Zeng, PROS, Houston, TX, United States, Jonas Rauch, Mohit Mahajan, Manu Chaudhary

In business-to-business pricing, organizations aim to maximize profits by developing optimized strategies based on historical transactions across a wide range of customers, products, and other characteristics. Each customer typically receives a unique price for a product through negotiation, making it essential to consider both customer-specific prices (CSP) and market prices (MP) in pricing decisions. This task has become increasingly difficult due to the presence of millions of transaction records and countless possible combinations of transaction features. Moreover, the data collection process in negotiation settings can be flawed. Business owners often record only winning transactions, omitting losses, which limits the data available.

Current pricing solutions often rely on probability-based models to compute win-rate elasticity. However, this approach fails entirely if loss data is not recorded and therefore unavailable for modeling. To address this issue, we propose a win-only logistic model with a policy-based optimal price. The policy-based optimal price is derived using CSP and MP estimates generated by a deep neural network and determines a target for price changes. This information serves as an optimal price constraint, passed to the win-only logistic model along with all learned feature embeddings. The win-only logistic model is a Bayesian logistic model that computes the probability of a win, conditional on only observed wins. Within this structure, the model parameters -- framed in terms of the optimal price -- become identifiable and estimable.

This win-only logistic model has been implemented in daily pricing decisions, demonstrating high credibility and delivering strong practical results.

3 - Robust Revenue Management Under Market Volatility

Ezgi Eren, PROS, Inc., Houston, TX, United States

In an era marked by unprecedented market volatility, adaptive pricing and revenue management strategies have become increasingly vital. This study revisits the data-driven methodology introduced by Eren et al. (2024), emphasizing its resilience under volatile demand conditions. Unlike traditional revenue management frameworks that rely on demand forecasting and mathematical optimization, this approach leverages heuristic techniques and machine learning models. Our simulation results demonstrate exceptional robustness to demand shocks, with revenue outcomes consistently within 1% of the theoretical optimum across a wide range of demand parameter variations. We offer in-depth insights into the methodology's implicit adaptivity, achieved without an explicit feedback loop. Finally, we discuss practical implications for implementing this strategy in demand-volatile industries such as air cargo and car rental.

4 - Enhancing Price Elasticity Estimation for Airline Dynamic Pricing via Kde-Based Sampling and Clustering

Laleh Kardar, PROS, Houston, TX, United States, Ravi Kumar

We address the challenge of estimating price elasticities of passenger demand in airline dynamic pricing when no-purchase data is unavailable. Building on a two-stage Poisson semi-parametric framework that combines machine learning with causal inference techniques, we propose two enhancements to improve robustness in sparse data settings. First, we apply a KDE-based sampling strategy to densify sparse regions of the feature space without imposing strong parametric assumptions. This improves model stability in underrepresented segments. Second, we introduce a top-down clustering approach that replaces rigid partitioning with data-driven groupings that respect operational constraints while capturing behavioral variation across segments. Empirical results on real airline transaction data show that these enhancements lead to more stable and differentiated elasticity estimates, supporting more effective and interpretable pricing decisions.

TB54

Building B Level 2 B216

Estimation and Optimization for Data-Driven Decisions

Invited Session

Revenue Management

Chair: Ke Li, INSEAD, Fontainebleau, France

Co-Chair: Spyros Zoumpoulis, INSEAD, Fontainebleau, France

1 - Online Inventory and Pricing with (Biased) Offline Data

Mengzhenyu Zhang, University College London, London, United Kingdom, Hyun-Soo Ahn, Zijin Zhang

With the growing amount of offline data, using the available offline data to facilitate online decision making becomes relevant to practice. A prominent challenge of using offline data, however, is that the true online data-generating process may differ unpredictably from the offline distribution. In this paper, we study the problem of how to use available and potentially biased offline data to support online decision making for two canonical decision problems: the dynamic newsvendor, where the demand tail distribution plays a key role in the decision, and dynamic pricing, where the point estimations are critical. We develop two online learning frameworks that incorporate offline data. The soft-switch algorithm automatically balances offline information and online feedback without requiring bias knowledge, while the hard-switch algorithm leverages known distributional shifts. Both learning algorithms achieve the optimal regret up to a logarithmic factor. We also derive explicit bounds showing how regret improvements scale with the size and bias of the offline data.

2 - From Small to Large: a Graph Convolutional Network Approach for Solving Assortment Optimization Problems

Guokai Li, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Pin GAO, Stefanus Jasin, Zizhuo Wang

Assortment optimization involves selecting a subset of substitutable products (subject to certain constraints) to maximize the expected revenue. It is a classic problem in revenue management and finds applications across various industries. However, the problem is usually NP-hard due to its combinatorial and non-linear nature. In this work, we explore how graph convolutional networks (GCNs) can be leveraged to efficiently solve constrained assortment optimization under the mixed multinomial logit choice model. We first develop a graph representation of the assortment problem, then train a GCN to learn the patterns of optimal assortments, and lastly propose two inference policies based on the GCN's output. Due to the GCN's inherent ability to generalize across inputs of varying sizes, we can use a GCN trained on small-scale instances to facilitate large-scale instances. Extensive numerical experiments demonstrate that given a GCN trained on small-scale instances (e.g., with 20 products), the

proposed policies can achieve superior performance (90%+ optimality) on large-scale instances (with up to 2,000 products) within seconds, which outperform existing heuristic policies in both performance and efficiency. Furthermore, we extend our framework to a model-free setting where the underlying choice model is unknown but transaction data is available. We also conduct numerical experiments to demonstrate the effectiveness and efficiency of our proposed policies in this setting.

3 - Beyond SVD: the F-SVD Revolution for Discrete Choice Estimation

Man Luo, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Guillermo Gallego

We present Filtered Singular Value Decomposition (*F-SVD*), a novel matrix decomposition framework that addresses the longstanding challenge of enforcing structural constraints in low-rank approximations. While conventional dimensionality reduction techniques like Singular Value Decomposition (SVD) and Non-negative Matrix Factorization (NMF) offer foundational solutions, they falter in critical applications like discrete choice modeling, where matrices must satisfy strict non-negativity, normalization, and exact-zero constraints. *F-SVD* resolves this by: (1) embedding a *structured mask* into the factorization process, and (2) enabling precise adherence to constraints while retaining the flexibility of SVD and the interpretability of NMF.

Theoretically, we establish convexity properties in alternating optimization steps that guarantee convergence to parsimonious solutions with errors matching or surpassing standard SVD. Empirically, *F-SVD* demonstrates three critical advantages:

- (i) *Exact reconstruction* of constrained entries (e.g., strict zeros in probability matrices),
- (ii) Great performance on real-world datasets – achieving superior results in 29/30 IRI consumer preference cases at “rank”-1 compared to a suite of parsimonious models, with the remaining case matching performance at “rank”-2,
- (iii) Computational tractability enabling high-dimensional inference.

Current work extends validation to synthetic matrices and industrial choice datasets, with full theoretical extensions and empirical results to be presented at the INFORMS Annual Meeting. By unifying constraint preservation with adaptive low-rank approximation, *F-SVD* establishes a new paradigm for structured matrix factorization, particularly in applications requiring simultaneous precision and interpretability – from consumer analytics to biomedical signal processing and operations decision systems.

4 - Decision-Focused Adjustments for the Predict-Then-Optimize Framework with Applications in Pricing

Guan Wang, Rotman School of Management, Toronto, ON, Canada, Michael Albert, Max Biggs, Ningyuan Chen

A common approach in data-driven decision-making is the predict-then-optimize (PTO) framework, where a decision-maker first estimates an unknown parameter from historical data and then uses this estimate to solve an optimization problem. While widely used for its simplicity and modularity, PTO can lead to suboptimal decisions when the estimation step does not account for the structure of the downstream optimization problem. We study a class of problems where the PTO decision introduces asymmetry in the objective with respect to the unknown parameter, creating a systematic preference for under- or over-estimation when the estimate is noisy. To address this, we present a data-driven post-estimation adjustment that improves PTO's decision quality while remaining practical and easy to implement. We show that when the objective satisfies a curvature condition based on the ratio of the third to the second derivative, the adjustment takes a simple closed-form expression. This applies to a broad range of pricing problems, including those with commonly used linear, log-linear, and power-law demand models. Under this condition, we establish theoretical guarantees that our adjustment uniformly outperforms standard PTO asymptotically, and we precisely characterize the resulting improvement. Additionally, we extend our framework to multi-parameter optimization settings and discuss applications where estimation bias is present. Numerical experiments demonstrate that our method consistently enhances revenue performance, particularly in small-sample regimes where estimation uncertainty is most pronounced, making it especially well-suited for pricing new products or those with limited price variation.

TB55

Building B Level 3 B314

Revenue Management and Platform Operations

Invited Session

Revenue Management

Chair: Yunduan Lin, The Chinese University of Hong Kong, Santa Clara, CA, N/A, United States

1 - An Large Language Model Agentic System for Outcome Prediction

Jiayi Wang, Stanford University, Stanford, CA, United States, Mohsen Bayati

We present an agentic system composed of multiple large language model (LLM) agents that automatically support outcome prediction. Our results demonstrate that this fully automated pipeline achieves predictive performance comparable to machine learning and human experts.

2 - Probabilistic Design in the Operation of Service Systems

Jiangze (Jay) Han, Columbia University, New York, NY, United States, Christopher Ryan, Mahesh Nagarajan, Xin Tong

I will discuss two papers addressing infinite-dimensional optimization problems over the space of probability distributions in the context of service system design.

The first paper explores optimal decomposition of a user distribution into sub-distributions to maximize operational targets. The second studies online optimization, where the decision maker sequentially determines a sequence of probability distributions to minimize cumulative loss. Examples include matching supply with unknown and dynamic demand in an online fashion, when supply is a spatial distribution over the service region.

To solve these problems, we propose novel gradient descent algorithms in the space of probability distributions, leveraging advances in optimal transport theory and Wasserstein gradient flow.

Our key results include:

1. In the first paper, our gradient-based algorithm finds a feasible decomposition of user distribution which also satisfies our proposed optimality condition, at a certain rate.
2. In the second paper, we establish static and dynamic regret bounds that match the optimal bounds in Euclidean space.

3 - Learning User Behavior in a Social Network Under Limited Data

Yunduan Lin, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Haoting Zhang, Zeyu Zheng, Zuo-Jun Shen

Accurately calibrating diffusion models in network studies is essential for linking theoretical results to real-world applications, such as predicting user behavior in new network structures and optimizing network operations. Effective calibration depends on connecting model parameters with available data. However, this task is complicated by two primary challenges: the dynamic nature of network diffusion and the limitations of available data. The diffusion process is complex, often resisting simple analytical representation, while data scarcity — both temporal (inability to capture the entire diffusion trajectory) and spatial (difficulty in obtaining network-level data) — can hinder model accuracy, leading to results that may not generalize well.

This work introduces a novel approach for calibrating diffusion models based on a single observed network instance, without requiring full data from the entire diffusion process. We use simulation to generate network-structured data under different parameter settings and minimize the discrepancy between observed and simulated

outcomes. We also analyze the performance of calibration under both full data and limited data schemes, characterizing the discrepancy caused by data scarcity. This approach provides a practical solution for calibrating network models under data limitations and has broad applications in real-world network studies.

4 - Mitigating Waste for Perishable Products Through Markdown

Jing Yu, Tsinghua University, Beijing, China, People's Republic of, Ganesh Janakiraman, Anyan Qi, Yongbo Xiao

We study a grocery retailer selling a perishable product with a two-period lifetime. To mitigate waste, the retailer employs markdown pricing for old products nearing expiration. Our model incorporates customer substitution behavior, where customers prioritize their preferred product type but substitute if unavailable, with unmet demand lost. Through this framework, we analyze the retailer's optimal markdown and inventory replenishment policies to balance waste reduction and revenue maximization.

5 - Order Stacking in on-Demand Delivery Platforms

James Scott, Duke University, Durham, NC, United States, Bora Keskin, Robert Swinney

Food delivery platforms routinely bundle orders together and assign the entire "stack" of orders to a single driver to deliver. We examine the impact of a restaurant's service region's geography on where within that region the stacks should be large or small.

TB56

Building B Level 3 B315

Information Acquisition in Digital Platforms

Invited Session

Auctions and Market Design

Chair: Ying Gao, UBC, Vancouver, BC, V6T0C2, Canada

1 - Consumer Profiling via Information Design

Itai Fainmesser, Johns Hopkins Carey Business School, Baltimore, MD, United States, Ruslan Momot, Andrea Galeotti

A platform uses purchase histories to profile consumers, form market segments, and disclose them to a seller. The seller tailors prices to these segments, generating new data that supports further profiling. We characterize the platform's ability to learn consumer valuations over time. We show that the platform cannot accurately learn consumers' valuations above the seller's optimal uniform price, but can below it. There is a complementarity between information design and other tools available to the platform to influence pricing—such as listing fees or proportional sales fees—for learning the valuations of high-value consumers.

2 - Naïve and Bayesian Learning with Misinformation Policies

James Siderius, Tuck School of Business at Dartmouth College, Hanover, NH, United States, Mohamed Mostagir

We study the effectiveness of two canonical interventions aimed at curbing the spread of misinformation—censorship and algorithmic regulation—through the lens of an economic model of information assimilation. Our framework accommodates heterogeneous agents, including both naïve DeGroot learners and more sophisticated Bayesian updaters who behave strategically. For naïve agents, optimal policies take a relatively intuitive form: removing content with low informational value and promoting exposure to more diverse viewpoints. However, we find that these same interventions are often ineffective—or even counterproductive—when applied to populations of Bayesian agents. In particular, censorship can exacerbate misinformation by distorting inference, and algorithmic regulation is only beneficial when the underlying level of misinformation is sufficiently low. Our results underscore the importance of accounting for agent sophistication when designing regulatory responses to misinformation and suggest limits to one-size-fits-all policy approaches in complex information environments.

3 - Screening Before and Auditing After: Information Design to Support Environmental Investment

Victoria Hang, Texas Tech University, Lubbock, TX, United States

This paper examines a dynamic principal-agent problem motivated by environmental regulation, where a regulator seeks to induce a firm's costly, non-contractible investment—such as adopting cleaner technology or mitigation measures—prior to the realization of uncertain environmental risk. After the investment stage, the firm also chooses an unobservable level of ongoing compliance effort. The regulator cannot commit to future enforcement policies but can influence behavior through the design of ex-ante signals (e.g., permitting thresholds, ESG ratings) and ex-post audits (e.g., inspections, compliance tests).

The regulator's objective is to maximize social welfare by inducing both appropriate investment and sufficient compliance effort. I compare two institutional mechanisms: signal design before investment and auditing after effort is chosen. Each mechanism generates different incentive structures and imposes different constraints on the regulator's ability to balance investment incentives, rent extraction, and effort distortion. When investment is relatively inexpensive, signal design is optimal as it efficiently screens agents and aligns investment decisions with social objectives. In contrast, when investment costs are high, auditing becomes preferable because it limits rent extraction while preserving incentives for compliance.

I characterize equilibrium behavior under each regime and identify the optimal information structure in both cases. Comparative statics show how the choice between screening and auditing depends on investment cost, the elasticity of compliance effort, and the distribution of environmental risk. These results inform the design of environmental planning tools and regulatory regimes that operate without commitment power.

4 - Optimal Recommendation with Costly Verification

Ying Gao, UBC, Vancouver, BC, Canada

Consumers who are targeted for recommendation by a designer (an advertiser or e-commerce platform) can often choose to accept, reject, or verify the recommendation, e.g., by reading reviews. We ask how optimal recommendation policies depend on the cost of verification. A recommender who wishes to persuade the buyer to purchase and can commit to their recommendation policy would prefer more costly verification. On the other hand, a recommender who cannot commit to their policy would prefer it to be easier for buyers to obtain information about the recommended product, as this circumvents their excessive temptation to falsely recommend lower-quality products.

TB57

Building B Level 3 B316

Selected Auctions and Market Design Papers from the EC 2025 Conference III

Award Session

Auctions and Market Design

Co-Chair: Rad Niazadeh, Chicago Booth School of Business, Chicago, IL, United States

Co-Chair: Christian Kroer, Columbia University, New York, NY, United States

1 - Credit Rating Design Under Adverse Selection

Qinzheng Li, National University of Singapore, Singapore, Singapore, Yifan Feng, Jussi Keppo

We study a strategic credit rating model where investors rely on credit rating agencies (CRAs) to assess firms' default probabilities. Firms seeking funding can solicit ratings by paying the CRA, who then optimizes a profit-maximizing rating policy based on firms' quality and solicitation decisions. We find that, in equilibrium, high-quality firms are more likely to solicit ratings and receive higher ratings compared to low-quality ones. Although solicited ratings are observed to be higher than

unsolicited ones, this discrepancy diminishes when accounting for firms' intrinsic qualities. Our model suggests that even if the CRA is purely self-interested, the reported discrepancy between unsolicited and solicited ratings in the existing literature could be primarily explained by adverse selection on the demand side of the credit rating market.

2 - Delegation with Costly Inspection

Suho Shini, University of Maryland, College Park, MD, United States, Mohammad Hajiaghayi, Piotr Krysta, Mohammad Mahdavi

We study the problem of delegated choice with inspection cost (DCIC), which is a variant of the delegated choice problem by Kleinberg and Kleinberg (EC'18) as well as an extension of the Pandora's box problem with nonobligatory inspection (PNOI) by Doval (JET'18). In our model, an agent may strategically misreport the proposed element's utility, unlike the standard delegated choice problem which assumes that the agent truthfully reports the utility for the proposed alternative. Thus, the principal needs to inspect the proposed element possibly along with other alternatives to maximize its own utility, given an exogenous cost of inspecting each element. Further, the delegation itself incurs a fixed cost, thus the principal can decide whether to delegate or not and inspect by herself. We show that DCIC indeed is a generalization of PNOI where the side information from a strategic agent is available at certain cost, implying its NP-hardness by Fu, Li, and Liu (STOC'23). In fact, we observe that DCIC becomes significantly more challenging than PNOI or the delegated choice problem in several aspects. First, neither of (i) running PNOI policy without delegation nor (ii) running a simple delegation mechanism can achieve a constant approximation to the optimal mechanism for DCIC, implying that we need to use both the delegation and inspection for efficient mechanisms. Furthermore, the standard approaches to PNOI (or Pandora's box problem) for upper bounding the optimal policy in a structured way to obtain algorithms, do not easily extend to DCIC.

3 - The Pseudo-Dimension of Contracts

Tomasz Ponitka, Tel Aviv University, Tel Aviv, Israel, Paul Duetting, Michal Feldman, Ermis Soumalias

Algorithmic contract design studies scenarios where a principal incentivizes an agent to exert effort on her behalf. In this work, we focus on settings where the agent's type is drawn from an unknown distribution, and formalize an offline learning framework for learning near-optimal contracts from sample agent types. A central tool in our analysis is the notion of pseudo-dimension from statistical learning theory. Beyond its role in establishing upper bounds on the sample complexity, pseudo-dimension measures the intrinsic complexity of a class of contracts, offering a new perspective on the tradeoffs between simplicity and optimality in contract design. Our main results provide essentially optimal tradeoffs between pseudo-dimension and representation error (defined as the loss in principal's utility) with respect to linear and bounded contracts. Using these tradeoffs, we derive sample- and time-efficient learning algorithms, and demonstrate their near-optimality by providing almost matching lower bounds on the sample complexity. Conversely, for unbounded contracts, we prove an impossibility result showing that no learning algorithm exists. Finally, we extend our techniques in three important ways. First, we provide refined pseudo-dimension and sample complexity guarantees for the combinatorial actions model, revealing a novel connection between the number of critical values and sample complexity. Second, we extend our results to menus of contracts and piecewise linear contracts, showing that their pseudo-dimension scales linearly with the menu size and polynomially with the number of linear pieces.

4 - Stable Matching with Contingent Priorities

Ignacio Rios, The University of Texas at Dallas, Richardson, TX, United States, Federico Bobbio, Margarida Carvalho, Alfredo Torrico

Using school choice as a motivating example, we introduce a stylized model of a many-to-one matching market where the clearinghouse seeks to implement contingent priorities—i.e., priorities that depend on the current assignment—to improve the likelihood that students with siblings are assigned together. We provide a series of guidelines and introduce two natural approaches to implement them: (i) absolute, whereby a prioritized student can displace any student without siblings assigned to the school, and (ii) partial, whereby prioritized students can only displace students that have a less favorable lottery than their priority provider. We study several properties of the corresponding mechanisms, including the existence of a stable assignment under \emph{contingent} priorities, the complexity of finding one if it exists, and its incentive properties. Furthermore, we introduce a soft version of these priorities to guarantee existence, and we provide mathematical programming formulations to find such stable matching or certify that one does not exist. Finally, using data from the Chilean school choice system, we show that our framework can significantly increase the number of students assigned to their top preference and the number of siblings assigned together relative to current practice.

5 - Admissibility of Completely Randomized Trials: A Large Deviation Approach

Chao Qin, Stanford University, Palo Alto, CA, United States, guido Imbens, Stefan Wager

When an experimenter has the option of running an adaptive trial, is it admissible to ignore this option and run a non-adaptive trial instead? We provide a negative answer to this question in the best-arm identification problem, where the experimenter aims to allocate measurement efforts judiciously to confidently deploy the most effective treatment arm. We find that, whenever there are at least three treatment arms, there exist simple adaptive designs that universally and strictly dominate non-adaptive completely randomized trials. This dominance is characterized by a notion called efficiency exponent, which quantifies a design's statistical efficiency when the experimental sample is large. Our analysis focuses on the class of batched arm elimination designs, which progressively eliminate underperforming arms at pre-specified batch intervals. We characterize simple sufficient conditions under which these designs universally and strictly dominate completely randomized trials. These results resolve the second open problem posed in Qin [2022]. The full version of this paper is available at <https://arxiv.org/pdf/2506.05329>.

TB58

Building B Level 3 B317

AI and Data in Healthcare

Contributed Session

Contributed

Chair: Jeremy Watts, I-O Urology, Chattanooga, TN, United States

1 - Addressing Rural Urban Health Disparities Through Predictive Medicine and AI-Driven Decision Support

Dursun Delen, Oklahoma State University, Tulsa, OK, United States, William Paiva

The disparity in access to healthcare between urban and rural communities remains a critical public health challenge, particularly due to the scarcity of subspecialty care in rural areas—termed here as the “lack of 'ologists.” This shortage places undue burden on rural primary care physicians who must manage complex cases without specialist support, exacerbating health disparities and compromising outcomes. This study explores the role of artificial intelligence (AI) in addressing these systemic gaps through the development and deployment of clinical decision support systems (CDSS). Specifically, we present a case study from Oklahoma State University's Center for Health Systems Innovation: a predictive AI tool designed to support diabetic retinopathy screening in rural primary care settings. This tool enables frontline physicians to make earlier and more accurate referrals, helping to prevent vision loss in diabetic patients. By leveraging real-world data and predictive modeling, such AI systems offer scalable, low-cost interventions to improve diagnostic accuracy and patient management in underserved regions. We discuss the challenges of implementation, including data quality, physician adoption, and workflow integration, and propose a framework for scaling AI-enabled tools in rural health ecosystems. This work highlights the promise of AI in mitigating rural health inequities and empowering local providers.

2 - Happily, Ever After &Ndash; the Effect of Marriage and Children on Life Expectancy in the United States

Banafsheh Behzad, California State University, Long Beach, Long Beach, CA, United States, Susan Martonosi, Katja Crusius

Life expectancy refers to the average number of years a person is expected to live based on various demographic and health factors. It is often used as a statistical measure to assess the overall health and mortality rates of a population. Life expectancy can vary significantly from one place to another and is influenced by factors such as health related factors, socioeconomic conditions, lifestyle choices, genetics, and more. This study examines the effect of marriage and children on life expectancy in the United States at the state level using data spanning from 2000 to 2017 in a fixed effects model. The findings suggest that, after controlling for other factors, states where a significant portion of the population is both married and employed tend to experience higher average life expectancies. Additionally, states with a substantial percentage of individuals having at least one child and having an annual income exceeding \$50,000 tend to exhibit increased average life expectancies among their populations.

3 - Mapping Perinatal Care Pathways to Address Maternal Health Disparities by Race

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

This study maps the care pathways of maternal patients during the perinatal period by race in order to improve the understanding of maternal disparities in morbidity and mortality for Black and White patients. We utilize electronic health record (EHR) data to model the care trajectory of patients during the perinatal period (i.e., prenatal, labor and delivery, and first year postpartum) as a function of race, age, socioeconomic status, insurance status, and comorbidities using continuous time stochastic modeling. Specifically, for each patient, we track each patient-provider interaction during pregnancy, birth, and postpartum. The proposed work seeks to use data analytics, text analytics and systems science methods to (i) explore the role of race in birth outcomes using the EHR, (ii) characterize the care trajectories for Black and White patients through the perinatal period, and (iii) identify opportunities for intervention ante-, intra-, and post-partum.

4 - Enhancing Diagnostic Accuracy in Noisy Breast Ultrasound Images Through Hybrid Cnn&Ndash;Transformer Denoising

Mariana Arriz Jorquiera, University of South Florida, Tampa, FL, United States, Ahtesham Bakht, Yashraj Jadhao, Ismail Uysal, Miguel Del Toro, Jose Zayas-Castro

Noise in medical imaging can compromise image clarity and significantly reduce the performance of diagnostic models. This is particularly evident in breast ultrasound, where variability in image acquisition and ultrasound-specific noise can make it difficult to detect important features. In this study, we examine the impact of three clinically relevant types of noise: Gaussian (sensor-related), Poisson (signal-dependent), and Speckle (coherent noise from ultrasound), each applied at low, moderate, and high intensity levels. Our findings show a consistent decline in classification accuracy as noise levels increase, highlighting the vulnerability of current machine learning models in realistic imaging conditions.

To address this challenge, we explore the application of a hybrid denoising model that combines a convolutional neural network (CNN) encoder, a transformer encoder, and a CNN decoder. Although CNN–transformer combinations have been applied in other domains, their use for denoising breast ultrasound images remains largely unexplored. By capturing both local spatial features and global contextual patterns, this architecture aims to reduce noise while preserving critical diagnostic information. We evaluate its performance by assessing improvements in classification accuracy on noisy and denoised images against established baselines, including Fourier-based methods and a conventional CNN denoiser. This study offers insight into how hybrid denoising strategies can improve diagnostic reliability in settings where noise is difficult to avoid.

5 - AI-Driven Assessment of BPH Symptoms from Multi-Day at-Home Urological Testing

Jeremy Watts, IO Urology, Knoxville, TN, United States, John Knoedler

We present **LUCI** (Lower Urinary Computational Insights), an AI-enabled system for objective evaluation of lower urinary tract symptoms (LUTS) associated with benign prostatic hyperplasia (BPH). LUCI analyzes time-series uroflow data collected by **CarePath** (iO Urology Corporation), a non-invasive, at-home testing platform that captures all voiding events over a multi-day period.

LUCI employs a modular AI architecture that integrates signal processing, anatomical estimation, and symptom modeling. From raw flow rate data, the system infers anatomical parameters and generates clinically aligned metrics, including the International Prostate Symptom Score (IPSS) and nocturnal polyuria frequency. This approach enables consistent, repeatable characterization of symptom severity—addressing the variability and subjectivity of traditional clinical assessments.

We evaluate LUCI on a dataset of BPH patients collected at Urology of Virginia. Results demonstrate LUCI's ability to replicate symptom trends captured by IPSS, quantify nocturnal voiding patterns, and provide continuous, real-world symptom monitoring. This work highlights the potential for AI-driven systems to enhance at-home diagnostics and support clinical decision-making in urology.

TB59

Building B Level 3 B318

Machine Learning for Healthcare

Contributed Session

Contributed

Chair: Yaoyao Huang, Simon Fraser University

1 - Predicting Permanent Molar Caries in Children and Adolescents Using Machine Learning

Lin Li, Kennesaw State University, Marietta, GA, United States, Pritam Deb, Christina Scherrer

Dental caries in permanent molars significantly impacts the well-being of children and adolescents, yet predictive modeling in this area remains limited. This study developed and tested machine learning (ML) models to predict caries using NHANES data from U.S. youth aged 6–19. We used the 2011–2014 cycles for model training/validation and 2015–2016 for testing. The outcome was a binary indicator of caries in at least one permanent molar, with predictors including demographics, diet, BMI, and behavioral factors.

Among 5,710 participants, 37.3% had molar caries. Of five ML models evaluated, logistic regression had the highest AUC (78.5%), the deep neural network (DNN) had the best accuracy (70.0%) and F1 score (64.9%), and the support vector machine (SVM) led in recall (81.3%). SHAP analysis of the XGBoost model highlighted age, family income, and household education as key predictors.

Our results show that ML models can effectively identify children at higher risk of molar caries, offering a potential tool for targeted screening and intervention strategies in oral health.

2 - Advanced Deep Learning for Predicting RNA Secondary Structure

You Zhou, Arizona State University, Tempe, AZ, United States, Giulia Pedrielli, Teresa Wu

RNA molecule plays a critical role in Gene Regulation, Catalysis, and Protein Synthesis. Nowadays, RNA has been applied in various biomedical applications, such as Genome Editing, Vaccines, and Cancer Immunotherapy.

Having a good understanding on how to predict RNA structure can benefit understanding the function of RNA. When predicting RNA structure, there are two challenges. Firstly, enumerating all feasible structures for an RNA sequence has searching space which increases exponentially and is intractable to resolve. Secondly, determining the RNA structure by using experimental approaches, such as X-ray, has high time and expense cost.

In this study, we propose a deep learning (DL) based approach to predict RNA secondary structure for a given RNA sequence. The proposed DL model relies on Transformer. We test our proposed model on two benchmark datasets, where our proposed approach presents superior performance compared other state-of-the-art approaches. Our proposed approach introduces how to leverage the attention mechanism to tackle RNA secondary structure prediction problem efficiently and accurately.

3 - Machine Learning&Ndash;Guided Early Discharge in Acute Myocardial Infarction: Equity, Efficiency, and Physician-Algorithm Concordance in a Microsimulation of Cadillac Risk Score Implementation

Emily Garcia, Purdue University, West Lafayette, IN, United States, Zachary J. Hass

As machine learning (ML)-based clinical decision support systems (CDSSs) become increasingly embedded in healthcare workflows, evaluating their downstream clinical and economic effects particularly in time-sensitive conditions like acute myocardial infarction (AMI) is critical. This study builds on prior work assessing the algorithmic

fairness of the Controlled Abciximab and Device Investigation for Lower Late Angioplasty Complications (CADILLAC) Risk Score, an ML-enabled CDSS designed to predict one-year mortality and major adverse cardiac events (MACE) following AMI hospitalization. While earlier analysis revealed modest but notable disparities in predictive performance across racial and insurance groups, the clinical and economic consequences of these discrepancies remain poorly understood. To address this gap, we use a microsimulation model informed by 4,364 distinct AMI patients from the MIMIC-IV database to estimate the impact of CADILLAC-guided early discharge decisions under varying levels of physician-algorithm concordance. Key outcomes include the cost implications of prolonged versus reduced length of stay, rates of 30-day readmission associated with MACE, and one-year mortality. The model also evaluates clinical decision-making dynamics—specifically, physician agreement or override of CADILLAC recommendations—and stratifies results by demographic subgroup to examine differential impacts. This study provides a rigorous framework for quantifying the downstream consequences of CDSS use in AMI care. By simulating both economic and clinical outcomes, it offers insights into how algorithm-informed decisions interact with clinical judgment and patient characteristics, with the goal of supporting more equitable and effective deployment of CDSSs in practice.

4 - Predicting Drug Shortages : A Machine Learning Approach

Vincent Hargaden, University College Dublin, Dublin, Ireland, Sean McGarraghy, Nikos Papakostas, Abu Shakil Ahmed, Ali Ala, Jamie MacMahon

This research develops an initial machine learning model for prediction of shortages in drug supply chains. It starts with acquiring publicly available drug shortages data from two countries (Belgium and Italy). Pre-processing steps include dimensionality reduction, label encoding and text pre-processing. Additionally, feature engineering is applied to the temporal features to capture seasonal trends. The machine learning approach used to develop the prediction model is Extreme Gradient Boosting (XGBoost) due to its regression and classification potential. The model is refined through hyperparameter tuning, feature selection and time-series cross-validation. The output from the model is a prediction when a particular drug will be unavailable. The results from this initial model underscore the importance of the availability of more extensive historical shortage data through reporting, larger sample sizes and diverse features for better prediction of drug shortages.

5 - Predicting Metabolic Dysfunction-Associated Steatotic Liver Disease with Machine Learning: Trinetx Database

Mary Ogidigben, Pennsylvania State University, State College, PA, United States, Paul Griffin, Soundar Kumara

Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD) affects approximately 25% of the U.S. adult population and is the most common chronic liver disease. The condition is asymptomatic in most individuals, but the progression of the disease leads to liver cirrhosis and is a leading indication for liver transplant. Although universal screening is not recommended, early detection is beneficial since lifestyle changes can control or reverse the condition. Several MASLD risk prediction models have been developed and often modeled with machine learning techniques. However, most previous methods have been developed for homogenous populations and there is no consensus on the best predictive factors to use. We focus on the importance of creating an algorithmically fair, rigorous, and reproducible prediction model for MASLD. With this in mind, we tested different machine learning algorithms and ultimately selected a logistic regression model with only 10 features due to its explainability, interpretability, and model performance. Our final logistic regression model had an AUROC of 0.836, accuracy of 77.6, and F1 score of 0.641.

6 - Fairness-Aware Machine Learning for Dental Caries Prediction

Pritam Deb, Kennesaw State University, Marietta, GA, United States, Lin Li, Christina Scherrer

While several studies have identified factors influencing dental caries, understanding caries in permanent molars among children and young adults still lacks clarity. This study aimed to develop a fairness-aware machine learning (ML) model to predict dental caries using publicly available National Health and Nutrition Examination Survey (NHANES) data from 2011 to 2016. We focused on the presence of caries (DMFT: decayed, missing, filled teeth) in permanent molars, considering demographics, dietary habits, body mass index (BMI), and lifestyle behaviors. Out of 5,710 participants aged 6 to 19, about 37.3% had caries. Fairness analysis highlighted important disparities in model accuracy: Non-Hispanic Black and Mexican American groups were identified correctly more often compared to Non-Hispanic Asian and Multi-Racial groups. Meanwhile, Non-Hispanic White participants had the lowest false-positive rates, and Mexican Americans were flagged for follow-up most frequently. To improve fairness, we applied standard bias mitigation techniques. This fairness-aware ML model can serve as an equitable screening tool to enhance oral health interventions for children and adolescents.

7 - Evaluating the Reliability of Online Drug Reviews: a Comparative Analysis of User-Reported and Official Adverse Drug Events

Yaoyao Huang, Simon Fraser University, Burnaby, BC, Canada, JM Goh, Aishwarya Shukla

Online drug reviews on online platforms is a major source of health information for patients and consumers seeking information about medications. While the online reviews on such platforms may be valuable, they carry the risk of disseminating misinformation, particularly regarding adverse drug reactions. Such misinformation can potentially result in harmful health outcomes. We compare adverse effects reported in user reviews from a popular online drug review platform to data from a database based on FDA-approved labels and clinical trials. We report our initial findings on overreported ADRs compared to underreported ones. Additionally, we report the findings from our clustering analysis where we identify two prominent perceptual patterns among user-reported adverse events.

TB60

Building B Level 3 B319

LLM for Business and Operations

Contributed Session

Contributed

Chair: Vicky Zhu, Babson College, Wellesley, MA, 02457, United States

1 - LLMs in Product Selection for Small E-Commerce

Yi Jie, The University of Hong Kong, Hong Kong, Hong Kong, Shan Huang

With the rapid advancement of AI and the need to reduce costs and improve efficiency in traditional experimental methods, large language models (LLMs) have shown significant potential in supporting causal inference tasks. We propose an LLM-based multi-agent causal inference system to replace traditional online randomized experiments through a collaborative framework consisting of reasoning induction, attribution analysis, and feedback correction. This system enables the automated discovery and validation of effective product strategies while addressing key challenges, such as activating the causal inference capabilities of LLMs, designing a collaborative division-of-labor model for multi-agent systems, and optimizing the structured input of experimental data. By evaluating the framework using historical WeChat experimental data, we demonstrate its ability to enhance decision-making processes with greater effectiveness and reliability.

2 - Generative AI-Driven Customer Experience Optimization

Nuo Xu, Wenzhou-Kean University, Wenzhou, China, People's Republic of, Xinci Chang, Jing Ma

This study develops an integrative framework to investigate how Generative Artificial Intelligence (GAI) shapes customer experience and purchase intention. Focusing on functional competence of GenAI, previous studies fail to address its emotional characteristics and its impact. Drawing on Stimulus-Organism-Response (SOR) theory and Value-Based Adoption Model (VAM), we conceptualize GAI's functional and emotional features as external stimuli that generate distinct utilitarian and hedonic value perceptions, which ultimately drive customer satisfaction and purchase intention. To test this framework, we adopt a mixed-methods approach comprising quantitative surveys, experimental designs, and semi-structured interviews, thereby offering cross-cultural insights into the multifaceted impact of enAI. By disentangling the functional and emotional dimensions of GenAI interactions—and accounting for both positive and negative influences—the study contributes to a deeper theoretical understanding of GenAI-driven consumer experience and offers practical guidance on developing customized GenAI-powered solutions that optimize user satisfaction and loyalty.

3 - An LLM-Driven Pipeline for Extracting EU Taxonomy Disclosures from Fossil Fuel Annual Reports

Jiahe Huang, University of Waterloo, Waterloo, ON, Canada, Stanko Dimitrov, Jangho Yang

The EU Taxonomy, adopted in 2020 as a pillar of the EU Sustainable Finance Action Plan, provides a common yardstick for assessing environmental sustainability and guiding green investment. Firms began reporting eligibility data for climate objectives in 2022, with alignment disclosures phased in through 2025. While designed to combat greenwashing and redirect capital toward net-zero targets, its behavioral impact remains largely untested. Given their critical role in global emissions and regulatory exposure, fossil-fuel companies offer a valuable lens through which to assess the Taxonomy's influence. To examine its impact on greenhouse gas trajectories and green investment behavior among major emitters, we develop an automated data extraction pipeline targeting unstructured annual and sustainability reports from over 20 fossil fuel companies operating in the EU across ten fiscal years (2015–2024). The pipeline begins with Mistral OCR, which converts each PDF into structured JSON format and archives the output in an SQL database to ensure traceability and consistency. We then fine-tune and benchmark three state-of-the-art generative language models—GPT-4o Mini, LLaMA-3.3-70B-Instruct, and DeepSeek-V3—for multi-task learning on a manually labeled dataset of sustainability disclosures. The best-performing model processes the JSON files and generates structured tables detailing CapEx, OpEx, turnover, R&D investment flows, and other green activities, annotated with EU Taxonomy alignment scores and progress indicators. This approach enhances extraction speed and accuracy, enabling systematic, longitudinal analysis of the Taxonomy's real-world impact. Beyond this study, the pipeline provides a scalable, adaptable framework for large-scale data extraction from unstructured corporate reporting in other regulatory and ESG domains.

4 - Llm-Based Multi-Robot Task Allocation and Path Planning for Heterogeneous Robots in Search and Rescue Operations

Kaushik Kannan, Michigan Technological University, Houghton, MI, United States, Jung Yun Bae

Urban Search and Rescue (USAR) missions in post-disaster environments demand rapid, adaptive task allocation and path planning for heterogeneous robot teams. While classical Multi-Robot Task Allocation (MRTA) methods exhibit limitations—including a lack of workload balance, rigid adherence to task requirements, and cross-scenario generalization—this work demonstrates that large language models (LLMs) with proper prompt engineering offer a transformative solution. We propose MTU-LLM, a structured prompt engineering for LLMs to generate optimized task allocation and path planning scripts. This approach reduces development time, ensures scalability and consistency across environments, and accelerates robot deployment. Evaluated on standard maps with varying team sizes and victim counts, MTU-LLM achieves significantly faster computation, improved workload balance, and comparable task fulfillment to classical methods. We further explore prompt engineering techniques to optimize LLM performance, revealing optimization strategies for different model characteristics.

5 - Bridging the Gap Between Business Users and Optimization: a Novel LLM-Powered Interactive Framework for Production Planning

Lennard Heuer, Korea Advanced Institute of Science & Technology (KAIST), Daejeon, Korea, Republic of

This presentation introduces an interactive framework powered by Large Language Models (LLMs), enabling business users—without coding or operations-research expertise—to build, refine, and validate sophisticated production-planning optimization models through natural language dialogue. Unlike most existing approaches for converting natural language into optimization models, which generate code in a single pass, our LLM-powered system supports incremental model construction, feasibility checks, simulation-based feedback, and collaborative edits. As a result, tasks typically handled by algorithm engineers are delegated to a hierarchical multi-agent architecture built on AutoGen. Dedicated agents manage (i) information retrieval, (ii) optimization-model generation and execution, (iii) simulation-model construction (for requests involving randomness), execution, and statistical result analysis, and (iv) visualization. Early evaluations show strong usability, suggesting the potential of this approach to effectively bridge the gap between non-technical business users and advanced optimization modeling.

6 - Chitchat with AI: Understand the Supply Chain Carbon Disclosure of Companies Worldwide Through Large Language Model

Vicky Zhu, Babson College, Wellesley, MA, United States, Haotian Hang, Yueyang Shen, Michelle Li, Jose Cruz

In the era of the rapidly growing economy, environmental sustainability is a crucial aspect of economic and social development. Maintaining climate responsibilities while generating financial success is challenging for companies to sustain their global portfolio. The Carbon Disclosure Project (CDP) dataset, being the world's largest, most comprehensive dataset for documenting companies' efforts on carbon disclosure, captures various responses from major companies across different industries and regions. Harmonizing these open-ended textual responses poses a heterogenic data challenge that demands novel language processing analytics to comprehensively derive actionable knowledge. Currently, there lack of an extensive framework to quantitatively evaluate the reliability of survey responses and quantify their alignment with companies' behavior. We study the CDP data using large language models (LLMs) that excel at effective summarization, instruction following, and semantic analysis. We propose a novel framework that leverages LLMs to assess carbon disclosures from the corporate supply chain. Our analyses demonstrate a correlated trend in the environmental awareness of different companies within certain industries. This framework is applicable to other longitudinal text-based business surveys.

TB61

Building B Level 4 B401

Knowledge Fusion for Spatial-Temporal Data Analytics

Invited Session

Quality, Statistics and Reliability

Chair: Shenghao Xia, Bowling Green State University, Bowling Green, OH, United States

Co-Chair: Fenglian Pan, The University of North Carolina at Charlotte, Charlotte, NC, United States

Co-Chair: Jian Liu, University of Arizona, Tucson, AZ, United States

1 - Adaptive Data Collection for Post-Disaster Rapid Damage Assessment

Mostafa Reisi, University of Florida, Gainesville, FL, United States, Mohammad Ilbeigi, Hojat Behrooz

Post-disaster rapid damage assessment is essential for successful emergency operations. However, existing data collection solutions for rapid damage assessment rely on static prior data and lack adaptability to new observations during assessment operations. This lack of adaptability limits the ability of emergency responses to systematically and effectively address unprecedented and dynamic situations in a data-driven manner. To address this research gap, our study proposes a novel adaptive data collection framework for post-disaster rapid damage assessment. The framework aims to maximize information gain in the shortest possible time by effectively targeting the most informative locations, while accounting for data collection costs—measured by total travel distance—and promoting sample diversity to prevent biased estimations. More specifically, the method consists of two main components: (1) an ordinal probabilistic gradient boosting model that forecasts the probabilities of damage severity levels for buildings affected by a disaster, and (2) an adaptive sampling mechanism that iteratively evaluates the forecasting model's performance and enhances it by identifying and incorporating the most informative samples. The adaptive sampling process is guided by a multi-objective optimization problem designed to reduce prediction uncertainty, minimize data collection costs, and maintain balanced sampling. The optimization problem is solved using the Non-Dominated Sorting Genetic Algorithm II. We empirically validated the proposed framework using historical damage assessment data from Hurricanes Michael and Dorian. The results demonstrated the method's ability to continuously improve the predictive power of the forecasting model by identifying the most informative target observations while minimizing data collection costs and preserving sample diversity.

2 - Natural Hypergradient Descent: Algorithm Design, Convergence Analysis and Parallel Implementation

Shancong Mou, University of Minnesota Twin Cities, Minneapolis, MN, United States, Deyi Kong, Zaiwei Chen, Shuzhong Zhang

Bi-level optimization has emerged as a powerful framework for modern machine learning applications such as meta-learning, data synthesis, and hyperparameter tuning. Hypergradient descent is commonly used to solve bi-level optimization problems, but computing the hypergradient—particularly the Hessian inverse of the inner problem, either explicitly or implicitly—poses significant computational challenges when the inner problem involves large-scale models. In this paper, we propose a simple and scalable method, Natural Hypergradient Descent (NHGD), which follows an optimize-and-approximate strategy. NHGD approximates the Hessian inverse using a running average of the inverse empirical Fisher information matrix (EFIM), computed directly from the gradients generated by each inner-loop SGD update. We provide a detailed non-asymptotic analysis demonstrating convergence to stationary points with iteration complexity matching that of existing optimize-then-approximate hypergradient methods, which first solve the inner problem to a desired accuracy and then approximate the Hessian inverse via Neumann series or conjugate gradient solvers. A key

advantage of NHGD is that its optimize-and-approximate design enables parallel computation of the hypergradient approximation alongside the inner-loop SGD steps. This leads to substantial improvements in computational efficiency, particularly for large-scale problems. Experiments on hyper data cleaning and learning data augmentation tasks demonstrate that NHGD achieves comparable or better performance with significantly reduced runtime overhead. Natural Hypergradient Descent: Algorithm Design, Convergence Analysis and Parallel Implementation

3 - Physics-Informed Attention-Enhanced Fourier Neural Operator for Solar Magnetic Field Extrapolations

Jinghao Cao, New Jersey Institute of Technology, Newark, NJ, United States, Qin Li, Mengnan Du, Haimin Wang, Bo Shen

We propose Physics-informed Attention-enhanced Fourier Neural Operator(PIANO) to solve the Nonlinear Force-Free Field (NLFFF) problem in solar physics. Unlike conventional approaches that rely on iterative numerical methods, our proposed PIANO directly learns the 3D magnetic field structure from 2D boundary conditions. Specifically, PIANO integrates Efficient Channel Attention (ECA) mechanisms with Dilated Convolutions (DC), which enhances the model's ability to capture multimodal input by prioritizing critical channels relevant to the magnetic field's variations. Furthermore, we apply physics-informed loss by enforcing the force-free and divergence-free conditions in the training process so that our prediction is consistent with underlying physics with high accuracy. Experimental results on the ISEE NLFFF dataset show that our PIANO not only outperforms state-of-the-art neural operators in terms of accuracy but also shows strong consistency with the physical characteristics of NLFFF data across magnetic fields reconstructed from various solar active regions.

4 - Modeling Spatially Correlated Failure-TIME Data Under Two Distance Functions with An Application to Titan Gpu Data

Jared Clark, Virginia Tech, Blacksburg, VA, United States, Jie Min, Yueyao Wang, Yili Hong, George Ostrouchov

One common approach to the statistical analysis of spatially correlated data relies on defining a correlation structure based solely on the physical distance between the locations of observed values. However, some data have a complex spatial structure that cannot be adequately described with the physical distance alone. In this work, the spatial failure-time data of focus contains information on GPUs that are linked through a series of wired connections, where it is expected that the failure-times of GPUs with few connections between them will be highly correlated. The proposed lifetime regression model includes random effects capturing the dependency due to physical location as well as random effects explaining the dependency due to the number of logical connections between GPUs. The analysis of this GPU dataset serves as an example of models with multiple spatial random effects and the ideas presented can be extended to other applications with complex spatial structures. A Bayesian modeling scheme is recommended for this class of analyses. The examples in this work use the software package, Stan, to produce Markov chain Monte Carlo draws for parameter estimation. This modeling effort is validated through simulation which demonstrates the accuracy of statistical inference. We also apply the developed framework to the large-scale Titan GPU failure-time data.

5 - Point-ITR: Task-Oriented Importance Sampling for Large-Scale 3D Point Clouds in Manufacturing

Yichen Ma, Georgia Tech, Atlanta, GA, United States, Michael Biehler, Chiehyeon Lim, Jianjun Shi

The increasing adoption of advanced three-dimensional (3D) scanning technologies has made large-scale point clouds containing millions of 3D measurement points standard in applications like manufacturing. However, processing immense amounts of 3D data imposes significant computational loads, often resulting in discarded critical information and suboptimal outcomes for downstream tasks. This paper introduces Point-ITR, a task-oriented sampling method tailored for regression tasks, which selectively retains the most informative points within large-scale point clouds. Specifically, we propose a gradient-based importance sampling framework for intra-sample selection (selecting points within a 3D point cloud) and a feature-based weighting scheme for inter-sample selection (selecting among different 3D point cloud sub-samples). Additionally, we introduce an iterative random sampling (ItrRS) module for preprocessing and an Offset Residual Block that utilizes a reference design model to learn structural features and accelerate both training and testing, which allows a simple fully connected network to process large-scale point clouds. Our approach improves prediction accuracy across downstream tasks while ensuring that the rich details captured are fully utilized for interpretation, offering a more effective and efficient solution. We validate our methodology through simulation studies and real-world case applications in additive manufacturing, demonstrating its robustness and practical applicability.

TB62

Building B Level 4 B402

Integration of Human, Knowledge and Systems for Quality and Personalized Decision-making

Invited Session

Quality, Statistics and Reliability

Chair: Feng (Ryan) Lin, University of Washington, Seattle, WA, United States

Co-Chair: Alan Vazquez, Tecnologico de Monterrey, Monterrey, 64700, Mexico

Co-Chair: Jiajing Huang, Kennesaw State University, Atlanta, GA, 30346, United States

1 - An Entropy-Based Causality Framework for Cross-Level Faults Diagnosis and Isolation in Building HVAC Systems

Jiajing Huang, Kennesaw State University, Atlanta, GA, United States

Faults, such as malfunctioning sensors, equipment, and control systems, significantly affect a building's performance. Automatic fault detection and diagnosis (AFDD) tools have shown great potential in improving building performances, including both energy efficiency and indoor environment quality. Since modern buildings have integrated systems where multiple subsystems and equipment are coupled, many faults in a building are cross-level faults, i.e., faults occurring in one component that trigger operational abnormalities in other subsystems. Compared with non-cross-level faults, it is more challenging to isolate the root cause of a cross-level faults due to the system coupling effects. Bayesian networks (BNs) have been studied for the root cause isolation for building faults. While promising, existing BN-based diagnosis methods highly rely on expert domain knowledge, which is time-consuming and labor expensive, especially for cross-level faults. To address this challenge, we propose an entropy-based causality learning framework, termed Eigen-Entropy Causal Learning (EECL), to learn BN structures. The proposed method is data-driven without the use of expert domain knowledge; it utilizes causal inference to determine the causal mechanisms between faults status and symptoms to construct the BN model. To demonstrate the effectiveness of the proposed framework, three fault test cases are used for evaluation in this study. Experimental results show that the BN constructed by the proposed framework is able to conduct building cross-level faults diagnosis with a comparable isolation accuracy to those by domain knowledge while maintaining less complexed BN structure.

2 - Human-Machine Collaborative Experimental Design for High-Quality Material Manufacturing

Guoyan Li, Northeastern University, Medford, MA, United States, Xiaoning Jin

Due to the substantial expenses associated with the manual annotation process, Bayesian Optimization (BayesOpt) is extensively employed within engineering optimization to substantially reduce the need for labor-intensive manual labeling through experimentation. BayesOpt involves a closed-loop where a probabilistic surrogate model is trained iteratively to guide sequential experimental design. Traditional BayesOpt methods that leverage a small labeled dataset to quantify the unlabeled information usually start a pure exploration of the whole design space in the initial optimization phase, which causes an over-exploration problem. In this work, we aim to improve the efficiency of exploring the entire design space by implementing a graph-based Gaussian Process as a surrogate model, which integrates geometric information within the design space. In addition, we designed a pseudo-label generator, which estimates the labels of informative yet unlabeled data, thereby expanding the training dataset. This new framework enhances the learning efficiency of the surrogate model, enabling more precise and faster design space exploration for optimal design. We

validated the effectiveness of our proposed approach through simulation data and applied this method to optimize mechanical structures in the additive manufacturing process.

3 - Reinforcement Learning Based Framework for Rna Secondary Structure Prediction

Menghan Liu, Arizona State University, Tempe, AZ, United States

Ribonucleic acid (RNA) is a fundamental biological molecule that is essential to all living organisms, performing a versatile array of cellular tasks. The function of many RNA molecules is strongly related to the structure it adopts. As a result, many efforts were put to solve the RNA secondary structure prediction problem: given a sequence of nucleotides, return a probable list of base pairs, referred to as the secondary structure prediction. Early algorithms have largely relied on finding the structure with minimum free energy. However, the predictions rely on effective simplified free energy models that may not correctly identify the correct structure as the one with the lowest free energy. In light of this, new, data-driven approaches that not only consider free energy, but also use machine learning techniques to learn motifs have also been investigated, and have recently been shown to outperform free energy based algorithms on several experimental data sets. In this work, we introduce the new ExpertRNA algorithm that provides a modular framework which can easily incorporate an arbitrary number of rewards and secondary structure prediction algorithms. We argue that this capability of ExpertRNA has the potential to balance out different strengths and weaknesses of state-of-the-art folding tools. We test the ExpertRNA on several RNA sequence-structure data sets, and we compare the performance of ExpertRNA against a state-of-the-art folding algorithm. We find that ExpertRNA produces, on average, more accurate predictions of non-pseudoknotted secondary structures than the structure prediction algorithm used, thus validating the promise of the approach.

4 - Bayesian Active Learning by Confidence Gradient-Based Mean Objective Cost of Uncertainty

Feng (Ryan) Lin, University of Washington, Mountlake Terrace, WA, United States

Many active learning methods adopt one-step-look-ahead acquisition functions to select the data to query their labels that can help improve the model performance the most. For example, expected loss reduction (ELR) or mean objective cost of uncertainty (MOCU) based methods focus on greedy maximization of uncertainty reduction related to learning objectives. However, without considering the long-term impact of the queried data, the one-step-look-ahead strategy suffers from myopic behavior. To relieve this myopia while still taking advantage of the simplicity and efficiency of one-step-look-ahead, we introduce a confidence cumulation mechanism by a new concept called confidence gradients (CG) and propose CG-MOCU to buffer against the risk of developing myopia for MOCU-based active learning. Theoretically, CG-MOCU will not get stuck in the long run and can converge to the true optimal classifier without additional constraints on the acquisition function. Empirical results on both synthetic and real-world data show that CG-MOCU outperforms the state-of-the-art active learning methods.

5 - Unsupervised Quality Control in Crowdsourcing via Mixture Random Utility Models

Hanming Zheng, City University of Hong Kong, Hong Kong, Hong Kong

Crowdsourcing is a transformative approach that harnesses collective intelligence to develop innovative and effective solutions. Despite its growing popularity, it faces challenges in ensuring the reliability and accuracy of crowdsourced results, making quality control crucial for effectively aggregating worker responses. This study proposes a novel probabilistic framework for worker response aggregation that unifies random utility model (RUM)-based choice modeling with latent worker clustering. We introduce a Mixture of Plackett-Luce Model (MPM), where worker responses are modeled as a mixture of choices from different latent clusters, each characterized by unique RUM parameters. The model parameters, including cluster-specific RUM attributes, task parameters, and worker cluster membership probabilities, are jointly learned using an expectation-maximization (EM) algorithm. This integrated approach enables mutual refinement between parameter estimation and cluster identification, leading to more accurate modeling. Simulation studies demonstrate the model's ability to recover ground-truth parameters, particularly the worker-task utility structure, across various data generation scenarios. Real-world case studies confirm that the proposed method outperforms existing methods in both accuracy and robustness. This work establishes a theoretically more coherent and powerful foundation for unsupervised quality control in crowdsourcing.

6 - Optimizing Two-Arm Clinical Trials for Personalized Medicine Using Integer Programming and Heuristic Algorithms

Alan Vazquez, Tecnologico de Monterrey, Monterrey, Mexico

Subjects with complex and serious diseases, such as malaria and leukemia, can have adverse reactions to the best-known treatments. The personalized medicine approach becomes beneficial because the subjects' characteristics or covariates are considered before a treatment assignment. Attractive statistical methods for personalized medicine determine the best treatment using the interaction between the treatments and the subject's covariates. To estimate this interaction, we use a two-arm clinical trial in which one of two treatments is assigned to each subject in a pre-selected sample. In this presentation, we introduce a statistical criterion to evaluate two-arm clinical trials in terms of the interaction estimation for all potential subjects' covariates, which can be quantitative or categorical. We develop theoretical results and two optimization algorithms to construct trials based on this criterion. Specifically, we introduce an integer programming algorithm that guarantees convergence to the optimal trial and a computationally efficient coordinate-exchange algorithm with a novel updating formula. Numerical experiments demonstrate the computational performance of our algorithms to construct two-arm clinical trials for various problems. As an application, we show that our algorithms can redesign a two-arm Scleroderma clinical trial with a smaller variance of the interaction estimate than the implemented trial for a test set of subjects.

TB63

Building B Level 4 B403

AI and Business Transformation

Contributed Session

Contributed

Chair: Wanxue Dong, The Chinese University of Hong Kong, Hong Kong, Hong Kong

1 - Digital Transformation in Traditional Manufacturing: a Three-Phase Strategy for Efficiency and Sustainability

Kun Zhong Lin, National Cheng Kung University/Department of Business Administration, Tainan, Taiwan, Tai-Yu Lin, Sheng-pin Wu, Chien-Chih Wang

Traditional manufacturers often struggle with outdated systems and poor tech integration in the face of globalization and digitalization. This study analyzes a Taiwanese marine diesel engine parts maker's digital journey through three phases: Digitization, Optimization, and Transformation. The company first adopted basic digital tools to streamline operations, then applied real-time data to improve monitoring and engagement, and finally visualized data to enhance decision-making. Using a case study approach and resource-based view, the study finds improved order speed, paperless processes, higher profits, and lower carbon emissions. A data-driven culture also boosted agility, communication, and leadership. Successful transformation aligns with core operations to enhance efficiency, competitiveness, and sustainability.

2 - AI Avatars on Douyin: Impacts on Engagement and Productivity

Yuzhou Chen, Southeast University, Nanjing, China, People's Republic of, Yulin Zhang, Yiting Guo, Haowen Fan

As generative artificial intelligence continues to reshape the creator economy on digital platforms, AI-generated avatars are emerging as powerful tools for influencers on video-sharing platforms. These avatars mimic an influencer's persona to engage with followers autonomously, raising important questions about their impact on creator performance. This study examines the causal effects of AI avatar adoption by leveraging a quasi-natural experiment on a major Chinese short-video platform. Combining difference-in-differences estimation with propensity score matching (PSM-DID), we assess how AI avatars influence two key outcomes: user engagement—measured by likes, comments, shares, and follower growth—and content output, measured by the number of videos posted. Our results show that AI avatars significantly increase user engagement across all metrics, but have no measurable impact on content production. This indicates that avatars improve the quality of interaction rather than stimulating greater creative output. Furthermore, moderation analysis reveals that the positive effects are more pronounced for utilitarian content, suggesting that content type moderates the effectiveness of AI-mediated engagement. These findings contribute to the literature on human-AI collaboration and influencer strategy by challenging the assumption that automation necessarily boosts productivity. Instead, they highlight AI avatars as strategic tools for deepening audience engagement without increasing labor demands. Practical implications for platforms and content creators include aligning AI avatar use with content type, audience expectations, and creator workload management.

3 - Content-Embedded Advertising Response Prediction for Informed Multimodal Branded Posts Diffusion: a Novel Deep Learning Method

Renzhi Gao, School of Management, Zhejiang University, Hangzhou, China, People's Republic of, Gang Chen, Xiangwei Kong, Junjie Wei

Content-embedded advertising is an emerging advertising format where sellers seamlessly integrate advertising content into multimodal social media posts, such as blogs, images, and videos (i.e., multimodal branded posts). These multimodal branded posts help make ads appear more natural and less intrusive. Diffusing multimodal branded posts through strategic traffic investment is effective operational strategy for sellers to boost user engagement and responses to their advertised products. However, when launching a new product, sellers often face challenges in cost-effectively creating and diffusing multimodal branded posts, including: (1) determining which topics to focus on during the post creation phase, (2) deciding how much traffic to invest in the post heating phase, and (3) allocating additional traffic to posts that have already received initial investment in the post diffusion phase. In this study, as inspired by the spreading activation theory, we propose a multimodal branded posts diffusion (MBPD) method, which: (1) simulates sellers' behavior in creating diversified branded posts across various topics in the post creation phase, (2) simulates sellers' traffic investment behavior across posts on certain topics in the post heating phase, (3) simulates sellers' additional traffic allocation behavior based on user responses to initial traffic investment in the post diffusion phase, and (4) optimizes the post diffusion framework through advertising response prediction. For a new product promotion, MBPD provides quantitative insights that help sellers optimize post creation, traffic investment, and additional traffic allocation for profitability. Empirical evaluation and explanatory analyses based on two real-world content-embedded advertising cases demonstrate the effectiveness of MBPD.

4 - Beyond Transparency: Explanations for Asymmetry and Irreversibility in AI-Assisted Decision Making

Nolan Talaei, University of Massachusetts Lowell, Lowell, MA, United States, Asil Oztekin, Hongwei Zhu, Luvai Motiwalla, Anol Bhattacharjee, Salih Tutun

Integrating artificial intelligence (AI) into decision-making and business processes, such as credit underwriting, pricing strategy, and operational planning, demands explainability beyond transparency to support targeted, domain-relevant action. Prevailing explainable AI (XAI) techniques often rely on static feature importances or one-shot counterfactuals, which overlook deeper behavioral properties of decision models. This work introduces a novel framework that leverages bidirectional perturbation analysis to expose asymmetric feature sensitivity and irreversibility, two properties critical to understanding and intervening in complex predictive systems.

By analyzing the effects of both increases and decreases in input features, the method uncovers cases where a shift in one direction significantly alters predictions while a change in the other direction has minimal or no impact. It further reveals decision irreversibility, where reversing an unfavorable outcome requires disproportionately large or qualitatively different input changes. Applied to a real-world consumer credit dataset characterized by binary lending outcomes and behavioral risk indicators, the approach identifies hidden thresholds, decision cliffs, fragile boundary regions, and zones of predictive inertia, offering plausible “what-if” scenarios aligned with feasible interventions.

Evaluations demonstrate that the method surfaces structural properties of model behavior that remain hidden under conventional explainability techniques. It consistently produces context-sensitive counterfactuals while maintaining interpretive stability across similar applicant profiles. These insights empower decision-makers to distinguish between fragile and stable cases, assess trade-offs more effectively, and intervene more confidently. This work reframes explainability as an interactive, scenario-driven tool for strategic reasoning in AI-assisted decision systems.

5 - Customized Assessment of Decision-Making Expertise Using Machine Learning

Wanxue Dong, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Dongcheng Zhang, Yunduan Lin, Weiwan Wang

Experts' decision accuracy is a fundamental determinant of professional outcomes such as compensation, retention, and task allocation, and it enables consumers to make more informed choices. While recent studies have introduced scalable and effective methods to evaluate overall decision accuracy, current approaches fall short in supporting customized, task-level assessments of experts' decision expertise. Given that decision quality often varies across different types of tasks, task-level evaluation is particularly valuable in high-stakes settings. In this work, we propose a novel and general framework for customized assessment of experts' decision quality, where the goal is to estimate how well an expert is likely to perform on specific decision tasks. To the best of our knowledge, this is the first study to address task-level evaluation of decision quality in a generalized, data-driven manner. We demonstrate the effectiveness of our method through a downstream application: assigning experts to new tasks based on their areas of expertise. Our approach achieves state-of-the-art performance, highlighting the practical value of customized evaluation of experts' decision-making quality.

TB64

Building B Level 4 B404

Special topics in Behavioral Operations Management

Invited Session

Behavioral Operations Management

Chair: Sebastian Villa, University of New Mexico, Albuquerque, NM, United States

Co-Chair: Vinit Tipnis, Carlson School of Management, University of Minnesota, Minneapolis, MN, United States

1 - Not-for-Profit Provider Operational Transparency and Donations

Kevin Mayo, Washington State University, Pullman, WA, United States, Telesilla Kotsi, Chengcheng Zhai

In many not-for-profit settings, providing beneficiary information is unethical or impossible. We examine the effects of provider transparency in the not-for-profit sector on the likelihood and amount of donations. We run experiments examining the value of provider information and different dimensions of operational transparency, information about reactive versus proactive programs, and information about direct aid versus overhead funding. We aim to first establish these results in laboratory conditions to identify the underlying mechanisms, then test our findings through real-world experiments.

2 - When Workers Monitor: How Engagement with Energy Management Systems Affects Energy Efficiency

Donggyu Jeon, Indiana University - Kelley School of Business, Bloomington, IN, United States, Christopher Chen, Owen Wu, Amrou Awaysheh

This study examines the impact of user engagement with energy management systems (EMS) on energy efficiency in manufacturing settings. EMS, a vital part of operational infrastructure, provides comprehensive data on energy and electricity usage, consumption flows and patterns, and CO2 emissions, collected by energy meters installed at manufacturing sites. Drawing on site-month data from a global conglomerate spanning 2017–2022, we examine how four dimensions of user engagement—frequency (sessions), breadth (users), volume (actions), and duration—affects EE. We find a significant positive relationship between user engagement with EMS and overall energy efficiency. Additionally, the nature of these engagement matters. Sessions exhibiting greater intensity and an explicit focus on energy management yield noticeably stronger improvements in energy performance than lower-intensity or administrative interactions. We further validate the causal effect of user engagement using an instrumental variable (IV) approach, where the number of energy meters installed—exogenously determined by corporate headquarters—serves as an instrument for user engagement. The results confirm that energy meters enhance EE only when site managers and operators actively exploit EMS energy data. We employ an inverse probability weighting to address potential selection bias in our IV design. Further, we conduct a control function approach with IV to account for any residual endogeneity bias in our IV estimations. Finally, we find that engineering teams and line workers—those in direct control of machinery and day-to-day processes—are especially effective at converting EMS insights into tangible energy savings. Collectively, our findings highlight the importance of human interventions in unlocking the full potential of EMS technology.

3 - Queuing and Donations: Effects of Waiting Time in Checkout Charity

Sebastian Villa, University of New Mexico, Albuquerque, NM, United States

Firms frequently give consumers the option to donate to a nonprofit organization (NPO) at checkout counters. This approach improves firm's image and allows NPOs to gain additional sources of donations. Through experiments, we investigate how operational and donation-related attributes influence consumers' donation at checkouts. Our findings have implications for optimizing charitable giving without compromising firms' customer satisfaction.

4 - The Impact of Scheduling Fairness on Employee Turnover

Paige Tsai, Harvard Business School, Boston, MA, United States, Ryan Buell

Using a combination of field and lab data, we explore drivers of job sustainability in the restaurant industry. We leverage shift data from a large restaurant group to better understand the impact of shift unfairness on employee turnover and find that relative unfairness — as compared to one's peers — is associated with higher employee turnover. A series of lab experiments suggests this relationship may be causal and reveals that perceived schedule fairness mediates the relationship between schedule quality and schedule ratings as well as likelihood to stay in the job. Taken together, these findings highlight the role of both absolute and relative working conditions in shaping employee retention. While previous research has emphasized the importance of sufficient, stable, and predictable scheduling on firm performance, our results indicate that perceptions of fairness with respect to these dynamics — shaped by relative differences in scheduling quality — could play a critical role as well.

5 - When Transparency Backfires: Why Contingent Discounts May Outperform Preannounced Ones

Federico Zimmerman, Harvard University, Cambridge, MA, United States, Joaquín Navajas, Gustavo Vulcano

We study how preannounced discounts influence consumer behavior when inventory is limited. We develop a two-period game-theoretic model in which rational consumers with identical valuations choose whether to buy at full price or wait for a possible markdown. We consider two settings: (i) discounts contingent on remaining inventory, and (ii) preannounced discounts. In the contingent case, the seller maximizes revenue by keeping the full price, prompting all consumers to buy early. In contrast, preannounced discounts yield both pure and mixed-strategy equilibria. In the pure strategy equilibrium, all consumers defer purchases, lowering seller revenue. In the mixed-strategy equilibrium, deeper discounts reduce deferral by increasing competition for limited inventory, thereby benefiting the seller. To test these predictions, we conducted an online experiment where participants chose to buy or wait under contingent and preannounced discount regimes. Contrary to rational expectations, deeper discounts increased the likelihood of deferral. Furthermore, consumers were more likely to wait when discounts were preannounced with certainty, suggesting a behavioral aversion to ambiguity. These tendencies led to lower effective prices and reduced retailer revenue, particularly at higher discount levels. Our findings reveal a disconnect between theoretical predictions and actual behavior. While preannounced discounts may mitigate strategic delay under rational models, they can encourage deferral in practice. This highlights the need to incorporate behavioral elements — such as ambiguity aversion — into pricing strategies to design more effective markdown policies and optimize revenue under inventory constraints.

TB65

Building B Level 4 B405

AI-Driven Entrepreneurship and Innovation

Invited Session

Artificial Intelligence

Chair: Tung Cu, Northeastern Illinois University, 5500 North St. Louis Avenue, Chicago, IL, United States

1 - Unlocking Competitive Advantage in the Age of Intelligent Systems

Dave Chatterjee, Duke University, Johns Creek, GA, United States

Artificial Intelligence is redefining the boundaries of what's possible in business. As the technology rapidly evolves, forward-thinking entrepreneurs are leveraging AI not just to enhance existing processes — but to reimagine entire industries. This keynote highlights the transformative role of AI in driving entrepreneurship and innovation, drawing from rich, real-world insights featured in the acclaimed Cybersecurity Readiness Podcast Series. Through the journeys of pioneering founders — from behavioral biometrics and insider threat detection to AI model security and deepfake fraud prevention — attendees will discover how AI is being used to build high-impact, scalable, and secure businesses. These innovation stories illuminate how personal adversity and professional expertise often intersect to spark ventures with global relevance. Yet, as AI enables unprecedented growth and agility, it also introduces new risks. This session explores the dual-use nature of AI — its power to enable as well as undermine — and presents pragmatic strategies for aligning AI innovation with ethical guardrails and cybersecurity resilience.

2 - Does Serendipity Mitigate Online Dating Burnout? a Randomized Field Experiment on Serendipity-Oriented Recommender Systems

Yue Guan, Communication University of China, Beijing, China, People's Republic of

Users of online dating platforms increasingly experience burnout. This is partly driven by recommender systems that reinforce sorting patterns based on explicit attributes such as social status, age, or geographic distance. Accordingly, we propose a serendipity-oriented recommender system by (1) relaxing constraints on explicit attributes while maintaining relevance through personalized preference learning, and (2) leveraging a deep learning attention network to extract intrinsic compatibility from profile Q&A texts. In a randomized field experiment on a leading online dating platform, users in the treatment group exhibited a 5.2% decrease in profile views and a 36% reduction in match requests, but achieved a 29.4% increase in match probability. Additionally, they showed a 38.5% increase in positive emotion scores, expanded their consideration sets across age and geography, and formed reciprocal matches more efficiently. Our findings offer causal evidence that incorporating serendipity into reciprocal recommendation can enhance match quality and mitigate user burnout.

3 - How Governance of Startup Ecosystems Can Unleash AI-Driven Innovations in Developing Countries

Lan Le, Metropolitan University Prague, Prague, Czech Republic, Tung Cu

Startup ecosystems consist of entrepreneurs, investors, academic institutions, government agencies, and support organizations that collaborate to foster innovation. Governance in these ecosystems involves regulatory policies, institutional support, financial incentives, infrastructure development, and cooperative frameworks that collectively enable startups to thrive. Effective governance can stimulate entrepreneurship, accelerate technology adoption, and contribute to sustainable economic development. This study investigates how governance of startup ecosystems can catalyze AI-driven innovations in developing countries. Specifically, it compares governance structures that support entrepreneurship leveraging generative AI (GenAI) in countries such as India and Vietnam, with reference to models from developed nations in Europe and the United States. The analysis focuses on both financial and non-financial support mechanisms, including the roles of public and private sector venture capital, technology training, social networks, and cultural factors. To address the research questions, data will be collected and analyzed across multiple dimensions of governance and ecosystem dynamics. The study follows a structured research process encompassing background and context, literature review, research questions, data collection and analysis, and conclusions with actionable recommendations. The goal is to develop strategic insights into optimizing governance frameworks for AI innovation in emerging markets.

TB66

Building B Level 4 B406

Adaptive Experimental Design And Policy Evaluation Under Nonstationarity

Invited Session

Data Mining

Chair: Kyra Gan, Cornell Tech, Cornell University, New York, NY, United States

Co-Chair: Jiamin Xu, Cornell University, New York, NY, 14850, United States

1 - Sample Complexity Reduction via Policy Difference Estimation in Reinforcement Learning

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

Abstract: An autonomous agent is placed into an unfamiliar environment with unknown rules and hidden rewards — how quickly can it learn to navigate and maximize its rewards? Efficient exploration is central to reinforcement learning (RL), with critical applications from robotics to personalized healthcare. In this talk, we focus on the pure exploration problem in contextual bandits and tabular RL, aiming to identify an approximately optimal policy from a set of candidates with high probability. Previous work in contextual bandits significantly improved sample efficiency by estimating only differences between policies rather than behaviors individually. However, existing methods in tabular RL fail to exploit this insight, instead estimating each policy's behavior directly. We investigate extending this efficient approach from bandits to RL and present a nuanced answer: difference estimation is sufficient for contextual bandits but fundamentally insufficient in tabular RL. Inspired by this limitation, we

propose anchoring policy estimates to a single reference policy, achieving substantial sample complexity improvements and delivering the tightest known bounds in tabular RL. Our findings provide new theoretical insights and practical strategies for efficient exploration in RL. Based on [joint work](#) with Adhyayan Narang, Andrew Wagenmaker, and Lillian Ratliff

2 - Efficient Learning for Finite-Horizon Restless Bandits

Kyra Gan, Cornell Tech, Cornell University, New York, NY, United States

In this work, we will present an efficient learning algorithm designed for finite horizon restless bandit problem. We adopt a budgeted thresholding contextual multi-armed bandit reformulation that we established in a prior work, and show that the reformulation can be efficiently learned through a novel construction of the ϵ -process in conjunction with a two-phase sampling algorithm.

3 - On the Limited Representational Power of Value Functions and Its Links to Statistical (in)Efficiency

David Cheikhi, Columbia University, New York, NY, United States, Daniel Russo

Identifying the trade-offs between model-based and model-free methods is a central question in reinforcement learning. Value-based methods offer substantial computational advantages and are sometimes just as statistically efficient as model-based methods. However, focusing on the core problem of policy evaluation, we show information about the transition dynamics may be impossible to represent in the space of value functions. We explore this through a series of case studies focused on structures that arises in many important problems. In several, there is no information loss and value-based methods are as statistically efficient as model based ones. In other closely-related examples, information loss is severe and value-based methods are severely outperformed. A deeper investigation points to the limitations of the representational power as the driver of the inefficiency, as opposed to failure in algorithm design.

TB68

Building B Level 4 B408

Advances in Machine Learning and Deep Learning for Multivariate Time Series Forecasting

Invited Session

Data Mining

Chair: Shouyi Wang, The University of Texas at Arlington, Arlington, TX, United States

Co-Chair: Dervis Ozay, The University of Texas at Arlington, Arlington, TX, United States

1 - Feature-Sparse Probabilistic Time Series Forecasting for Demand Prediction: Benchmarking and Advancing State-of-the-Art Methods

Dervis Ozay, The University of Texas at Arlington, Arlington, TX, United States

Accurate demand forecasting underpins key operational decisions in dynamic pricing, inventory planning, and marketing optimization. Traditional time series models such as ARIMA, VAR, and tree-based machine learning algorithms (e.g., Random Forest, XGBoost) often struggle to address modern forecasting challenges, including high-dimensional input spaces with redundant features, non-linear dependencies, sudden structural shifts, and uncertainty quantification in volatile markets. This study presents a twofold contribution. First, we conduct a comprehensive benchmarking of state-of-the-art (SOTA) forecasting approaches, including classical statistical models, machine learning algorithms, and deep learning architectures, assessing their predictive accuracy and robustness across multiple demand forecasting scenarios. Second, we introduce a **Feature-Sparse Probabilistic Demand Learning Model (FS-PDLM)**, a novel Transformer-based framework that jointly performs time series forecasting and embedded feature selection. By enforcing sparsity constraints, FS-PDLM identifies the most informative factors driving demand while quantifying forecast uncertainty. Experimental results on synthetic and real-world datasets demonstrate that FS-PDLM not only enhances prediction accuracy and interpretability but also maintains competitive or superior performance compared to leading SOTA models.

2 - Topology-Aware Conformal Prediction for Stream Networks

Jifan Zhang, Northwestern University, Evanston, IL, United States

Stream networks, a unique class of spatiotemporal graphs, exhibit complex directional flow constraints and evolving dependencies, making uncertainty quantification a critical yet challenging task. Traditional conformal prediction methods struggle in this setting due to the need for joint predictions across multiple interdependent locations and the intricate spatio-temporal dependencies inherent in stream networks. Existing approaches either neglect dependencies, leading to overly conservative predictions, or rely solely on data-driven estimations, failing to capture the rich topological structure of the network. To address these challenges, we propose Spatio-Temporal Adaptive Conformal Inference (STACI), a novel framework that integrates network topology and temporal dynamics into the conformal prediction framework. STACI introduces a topology-aware nonconformity score that respects directional flow constraints and dynamically adjusts prediction sets to account for temporal distributional shifts. We provide theoretical guarantees on the validity of our approach and demonstrate its superior performance on both synthetic and real-world datasets. Our results show that STACI effectively balances prediction efficiency and coverage, outperforming existing conformal prediction methods for stream networks.

3 - Spatio-Temporal Generative Foundation Model for Power Outage Prediction and Simulation

Feng Qiu, Argonne National Lab, Lemont, IL, United States

As extreme weather events grow more frequent and grid topologies more complex, electric utilities require reliable tools to forecast where and when outages will occur and to simulate plausible outage scenarios for resilience assessment. Existing approaches for power outage analysis are typically tailored to a specific machine learning task or local service territory, which limits their scalability and transferability. Inspired by the success of large-scale generative pretraining in language and vision, we propose the first spatio-temporal generative foundation model for power outages that learns from massive archives of customer outage records across the U.S. alongside multi-modal covariates (e.g., weather, census data, and network topology). Our model employs a flow matching objective within a latent generative framework, ensuring generated trajectories remain statistically consistent with observed data. Inference then reduces to sampling from the model's conditional distributions, unifying prediction and simulation under a single probabilistic umbrella. The learned outage dynamics can be transferred to region-specific data with minimal fine-tuning effort. Preliminary experiments demonstrate that our model delivers strong predictive accuracy for outage forecasting and generates realistic scenarios for "what-if" resilience analyses. This equips utilities and planners with a flexible, data-efficient tool to anticipate emerging risks and stress-test grid resilience under diverse conditions.

4 - Change-Point Detection in Time Series Using Mixed Integer Programming

Artem Prokhorov, University of Sydney Business School, CEBA, CIREQ, Sydney, Australia

We use cutting-edge mixed integer optimization (MIO) methods to develop a framework for detection and estimation of structural breaks in time series regression models. The framework is constructed based on the least squares problem subject to a penalty on the number of breakpoints. We restate the L_0 -penalized regression problem as a quadratic programming problem with integer- and real-valued arguments and show that MIO is capable of finding provably optimal solutions using a well-known optimization solver. Compared to the popular L_1 -penalized regression (LASSO) and other classical methods, the MIO framework permits simultaneous estimation of the number and location of structural breaks as well as regression coefficients, while accommodating the option of specifying a given or minimal number of breaks. We derive the asymptotic properties of the estimator and demonstrate its effectiveness through extensive numerical experiments, confirming a more accurate estimation of multiple breaks as compared to popular non-MIO alternatives. Two empirical examples demonstrate usefulness of the framework in applications from business and economic statistics.

TB69

Building B Level 4 B409

Resilient and Optimized Rail Freight Networks

Contributed Session

Contributed

Chair: Yiduo Huang, University of Illinois Urbana Champaign, 103 E Healey St, Champaign, IL, 61820, United States

1 - How Intercontinental Rail Freight Planning Responses to Geopolitical Risks**Jing Shan, Chair of Logistics Management, Zürich, Switzerland, Stephan Wagner**

Global supply chains are becoming more volatile as geopolitical conflicts interrupt transport flow patterns and require constant redesign of transport networks. In particular, the Russia–Ukraine conflict has pushed shippers and planners to seek safer alternatives. The Middle Corridor has rapidly emerged as a pivotal rail link between Asia and Europe, offering a strategically important bypass that avoids Russian territory. A significant amount of financial resources has been reallocated to the Middle Corridor. However, it remains questionable to what extent the investment could improve throughput, decrease average delay per container, reduce average cost, and increase average profit and margins. Transport planners and policymakers need an effective tool capable of analyzing potential geopolitical disruptions and adapting network designs accordingly. However, existing rail planning research mostly neglects geopolitical factors. This paper therefore introduces the Geopolitically Intercontinental Service-Network Design (G-ISSND) model. The G-ISSND model integrates the routing preferences of avoiding Russia territory, while also considering service quality of the supply chain, rail resource efficiency, rail network capacity constraints, and border crossing times. Our results demonstrate that the performance of an intercontinental rail network must be optimized by balancing capacity expansion with faster transit times, while also considering the time-sensitivity of the goods structure and routing preferences of the shippers. The G-ISSND model serves as both a strategic and operational planning tool, enabling policymakers and transport planners to better respond to geopolitical disruptions in intercontinental rail freight network.

2 - Strategic Resilience in Eurasian Trade Corridors: a Game-Theoretic Approach**Nahid Jafari, SUNY, Farmingdale, Farmingdale, NY, United States, Bruno Kamdem**

Ensuring resilient and sustainable supply chains for goods and energy between Europe and Asia presents a complex challenge, influenced by long distances, political risks, economic sanctions, environmental concerns, and volatile geoeconomics conditions. This study focuses on a transnational transport corridor that spans multiple jurisdictions, each with distinct economic and geopolitical agendas. The region is further complicated by fluctuating commodity prices driven by geopolitical tensions and evolving international agreements.

To address these challenges, we develop a model based on Stochastic Differential Game Theory and Optimal Control, capturing the strategic interactions among key stakeholders, including transportation companies and national governments. In our framework, transportation firms (Player 1) aim to maximize profit while minimizing operational costs, including taxes, by leasing infrastructure—such as roads, railways, or pipelines—from corridor authorities controlled by governments (Player 0), who seek to maximize revenue from taxes and commodity sales.

These interactions unfold within a stochastic game setting, where the market value of a unit of the transported commodity is modeled as a log-normal mean-reverting process, reflecting both price volatility and long-term stabilization tendencies. This approach enables the formulation of robust, adaptive strategies that integrate transportation costs and geopolitical risks. We identify a unique Nash Equilibrium that optimally balances taxation and transport rates, thereby minimizing disruptions and fostering economic stability across politically sensitive and economically volatile regions.

3 - Practice-Based Optimization for Strategic Locomotive Assignment Problem**Yunji Kim, Georgia Technology of Institute, Atlanta, GA, United States, Amira Hijazi, Kevin Dalmeijer, Pascal Van Hentenryck**

This study presents a strategic optimization framework for the Locomotive Assignment Problem (LAP), developed in collaboration with a major North American Class I Freight Railroad. The problem is formulated as a network-based integer program over a cyclic space-time network, producing a repeatable weekly assignment plan for locomotives. The model incorporates a comprehensive set of real-world operational constraints while endogenizing the placement of work events—i.e., locomotive pick-ups and set-outs—throughout space and time. Unlike prior studies that treat these locations as fixed, this approach jointly optimizes them, improving feasibility and effectiveness of subsequent planning. One of the key contributions of this work is solving the LAP exactly with all feasible light travel arcs, a task previously considered computationally intractable due to the sheer scale of the arc set. To overcome this, a set of network-topology-aware reduction rules are introduced that dramatically reduce the candidate arcs without compromising optimality. This enables exact solutions on large-scale instances for the first time and provides provable lower bounds for evaluating the true quality of heuristics methods. Extensive computational studies demonstrate how strategic decisions and network behaviors adapt under varying cost and capacity regimes. The results highlight key operational trade-offs and identify critical network components essential for robust system design. Beyond delivering scalable, high-quality solutions, the proposed framework serves as a practical decision-support tool grounded in the operational realities of modern freight railroads.

4 - Trailer Path Optimization on Large Scale Network Using Scheduled Services**Mohammed Faisal Ahmed, Georgia Institute of Technology, Atlanta, GA, United States, Ahmed El-Nashar, Pascal Van Hentenryck**

This paper introduces the trailer path optimization with schedule services problem (TPOSSP) and proposes a column generation-based heuristic (CG-Heuristic) to efficiently solve large instances in tactical planning settings. The TPOSSP aims at determining trailer routes over a time-dependent network using existing scheduled services, while considering tractor capacity constraints and time windows for trailer pickups and deliveries. The objective is to minimize both the number of schedules used and the total miles traveled. Both arc-based and path-based formulations of TPOSSP are examined. To improve computational efficiency, a novel pruning algorithm called Earliest Latest Start Time Pruning (ELSP) is introduced to identify the exact feasible search space for each trailer. A fast time dependent shortest path algorithm is developed to solve the pricing problem of column generation. To stabilize the column generation process, a weighted Dantzig-Wolfe decomposition is applied. The models and algorithms are evaluated using a real-world dataset from a major logistics service provider. ELSP achieves runtime reductions of 85%–99% for arc-based formulations and 53%–85% for the CG-Heuristic. The CG-Heuristic produces solutions within a 3% optimality gap for small instances and within a 5% gap for large instances. Additionally, the heuristic reduces the net cost of operations by up to 3% for large-scale scenarios.

5 - Integrated Optimization of Skip-Stop Freight Trains Timetabling and Delivery Routing on Multi-Line Urban Metro Systems**Younju Jung, Korea University, Seoul, Korea, Republic of, Haneul Kim, Soon-Kyo Lee, Taesu Cheong**

The lack of urban logistics hubs sparks interest in incorporating skip-stop freight trains into passenger train schedules. This approach aims to maintain punctuality for passenger trains while ensuring timely delivery of freight. However, existing research remains on single-line scenarios and neglects transfers, limiting its applicability to real-world rail networks. In this study, we propose a mixed-integer linear programming (MILP) model that determines the timetables for skip-stop freight trains and the routes for freight delivery on a multi-line rail network at the same time. In order to solve large-scale instances efficiently, three methods are applied: (i) reducing the network based on commodities, (ii) determining the maximum number of freight trains per interval, and (iii) using a column generation-based reformulation. The effectiveness of the proposed method is verified through computational experiments and a case study on the Seoul Metro network. Furthermore, the unit prices per order that reflect target margins and the number of freight trains required to meet demand forecast are analyzed, offering valuable insights for strategic decision-making.

6 - Multiscale Data-Driven Intermodal Freight Traffic Assignment on Large Networks**Yiduo Huang, University of Illinois Urbana Champaign, Champaign, IL, United States, Yanfeng Ouyang, Jesus Osorio**

Intermodal freight transportation seamlessly integrates modes such as road, rail, and water to reduce congestion, costs, and environmental impacts. However, modeling these systems at a national scale is challenging due to heterogeneous data sources, varying regional shipper and carrier decisions, unknown regional and inter-modal transfer delay functions, and significant computational demands. To address these issues, this paper proposes a multi-source, data-driven intermodal freight traffic assignment framework that combines high-resolution regional data and simulation outputs with lower-resolution, large-scale network information to predict overall flow distributions and delays. Our hierarchical hybrid modeling approach represents each regional sub-network as a hyperlink in a hypergraph. We use an integrated inverse optimization and regression model with kernel methods to learn hyperlink travel time functions from both large-scale and regional data, and then apply a gradient-projection algorithm to solve for equilibrium flows on the resulting hypergraph. We also establish generalization guarantees for the learned kernel-based hyperlink cost

function, showing how the quantity of both upper- and lower-level data influences the expected prediction error. This integrated methodology enables robust evaluations of intermodal configurations and operational strategies, ultimately promoting more efficient and sustainable freight systems.

Tuesday, October 28, 1:15 PM - 2:30 PM

TC01

Building A Level 3 A301

JMP/Gurobi

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Bayesian Optimization for Efficient Decision-Making in Complex Systems

Ross Metusalem, JMP Statistical Discovery, Tampa, FL, United States

Bayesian Optimization (BO) is a sequential learning method for efficiently optimizing noisy, black-box, or expensive-to-evaluate functions. Starting with a small set of observations, BO fits a flexible probabilistic model to the data and uses that model to determine the next observation(s) to collect. Iterating on this process, BO balances exploration of uncertain areas of the factor space with exploitation of potential optima to arrive at a globally optimal solution. This session will establish the foundational concepts underlying BO and present a case study of BO in action.

2 - Optimization Education with Gurobi: A Social-Impact Case Study

Lindsay Montanari, Gurobi Optimization, Minneapolis, MN, United States, Everett Dutton

Gurobi, in partnership with ESUPS (Emergency Supply Pre-positioning Strategy), presents a new educational case study showcasing ESUPS' utilization of optimization to stage relief inventory for humanitarian organizations. ESUPS unites NGOs, UN agencies, governments, and academics to improve logistics preparedness. The STOCKHOLM platform (STOCK of Humanitarian Organizations Logistics Mapping) applies data analytics techniques including optimization to determine where to pre-position supplies, minimize response times, and maximize coverage in crises where every hour matters. This case study guides students and instructors through the full optimization pipeline: translating a complex humanitarian challenge into data, math, and models that support real-world decision making. Freely available, it gives learners hands-on experience with optimization in action, demonstrating how these methods can transform disaster response and deliver measurable social impact.

TC02

Building A Level 3 A302

ICS Awards Session

Award Session

Computing Society

Chair: Thiago Serra, University of Iowa, Iowa City, IA, United States

Co-Chair: Merve Bodur, University of Edinburgh, James Clerk Maxwell Building, Edinburgh, EH9 3FD, United Kingdom

1 - Robust Paths: Geometry and Computation (ICS Student Paper Award Winner)

Hao Hao, Carnegie Mellon University, Pittsburgh, PA, United States, Peter Zhang

Applying robust optimization often requires selecting an appropriate uncertainty set both in shape and size, a choice that directly affects the trade-off between average-case and worst-case performances. In practice, this calibration is usually done via trial-and-error: solving the robust optimization problem many times with different uncertainty set shapes and sizes, and examining their performance trade-off. This process is computationally expensive and ad hoc. In this work, we take a principled approach to study this issue for robust optimization problems with linear objective functions, convex feasible regions, and convex uncertainty sets. We introduce and study what we define as the robust path: a set of robust solutions obtained by varying the uncertainty set's parameters. Our central geometric insight is that a robust path can be characterized as a Bregman projection of a curve (whose geometry is defined by the uncertainty set) onto the feasible region. This leads to a surprising discovery that the robust path can be approximated via the trajectories of standard optimization algorithms, such as the proximal point method, of the deterministic counterpart problem. We give a sharp approximation error bound and show that it depends on the geometry of the feasible region and the uncertainty set. We also illustrate two special cases where the approximation error is zero: the feasible region is polyhedrally monotone, or the feasible region and the uncertainty set follow a dual relationship. We demonstrate the practical impact of this approach in two settings: portfolio optimization and adversarial deep learning.

2 - Improving the Security of United States Elections with Robust Optimization (Harvey J. Greenberg Research Award Winner)

Bradley Sturt, University of Illinois Chicago, Chicago, IL, United States, Braden Crimmins

For more than a century, election officials across the United States have inspected voting machines before elections using a procedure called logic and accuracy testing (LAT). This procedure consists of election officials casting a test deck of ballots into each voting machine and confirming the machine produces the expected vote total for each candidate. We bring a scientific perspective to LAT by introducing the first formal approach to designing test decks with rigorous security guarantees. Specifically, our approach employs robust optimization to find test decks that are guaranteed to detect any voting machine misconfiguration that would cause votes to be swapped across candidates. Of all the test decks with this security guarantee, our robust optimization problem yields the test deck with the minimum number of ballots, thereby minimizing implementation costs for election officials. To facilitate deployment at scale, we develop a practically efficient exact algorithm for solving our robust optimization problems based on the cutting plane method. In partnership with the Michigan Bureau of Elections, we retrospectively applied our approach to all 6,928 ballot styles from Michigan's November 2022 general election; this retrospective study reveals that the test decks with rigorous security guarantees obtained by our approach require, on average, only 1.2% more ballots than current practice. Our approach has since been piloted in real-world elections by the Michigan Bureau of Elections as a low-cost way to improve election security and increase public trust in democratic institutions.

3 - Modeling and Solving Mixed-Integer Quadratic Optimization Problems (Ics Prize - Co-Winner)

Alper Atamturk, University of California - Berkeley, Berkeley, CA, United States, Andres Gomez, Shaoning Han

Quadratic optimization with discrete decisions lies at the core of modern optimization, connecting classical problems in operations research with pressing challenges in machine learning and signal processing. Yet the combination of nonconvex quadratic structure and combinatorial constraints makes these problems notoriously difficult. In this talk, I will present recent advances in mixed-integer nonlinear optimization that demonstrate how structural insights can be systematically leveraged to overcome these challenges.

The key idea is to identify and exploit simpler building blocks within complex models, such as low-dimensional sets, rank-one quadratic forms, or submodular nonlinear functions. For these structures, we can derive ideal convex relaxations that capture their combinatorial and algebraic properties exactly. Embedding these relaxations into mixed-integer optimization formulations yields models that are not only theoretically tight but also computationally effective.

This perspective leads to substantial improvements in solving small- and medium-scale problems directly with modern solvers, while also enabling tailored presolve routines and specialized algorithms that extend tractability to high-dimensional structured instances. Taken together, these results suggest a unifying approach: by decomposing nonlinear models into their convexifiable components, we can design formulations and algorithms that achieve both provable strength and practical efficiency.

4 - Acceleration by Stepsize Hedging (ICS Prize Co-Winner)

Jason Altschuler, University of Pennsylvania, Philadelphia, PA, United States, Pablo Parrilo

Can we accelerate the convergence of gradient descent without changing the algorithm — just by optimizing stepsizes? Surprisingly, we show that the answer is yes. Our proposed Silver Stepsize Schedule optimizes strongly convex functions in $k^{\log_2 p} = k^{0.7864}$ iterations, where $p=1+\sqrt{2}$ is the silver ratio and k is the condition number. This is intermediate between the textbook unaccelerated rate k and the accelerated rate $\sqrt{k}$ due to Nesterov in 1983. The non-strongly convex setting is conceptually identical and leads to an analogously accelerated rate $\epsilon^{-\log_2 p} = \epsilon^{-0.7864}$. We conjecture and provide partial evidence that these rates are optimal among all possible stepsize schedules.

The Silver Stepsize Schedule is an explicit non-monotonic fractal. Why should such stepsizes help? The core intuition is “hedging” between individually suboptimal strategies — short steps and long steps — since bad cases for the former are good cases for the latter, and vice versa. Properly combining these stepsizes yields faster convergence due to the misalignment of worst-case functions. This talk is based on a line of work with Pablo Parrilo that originates from my 2018 Master’s Thesis — which established for the first time that judiciously chosen stepsizes can enable accelerated convex optimization. Prior to this thesis, the only such result was for the special case of quadratics, due to Young in 1953.

TC03

Building A Level 3 A303

From Expertise to Influence: Advocacy Training for the INFORMS Community and the 2025 AGC Award Winner

Invited Session

Advocacy

Chair: Jeffrey M. Cohen, INFORMS, 5521 Research Park Drive, Baltimore, MD, 21228, United States

1 - From Expertise to Influence: Advocacy Training for the INFORMS Community

Jeffrey M. Cohen, INFORMS, Baltimore, MD, United States

Your voice matters—and when informed by data and expertise, it can shape the future of federal science and technology policy. Led by INFORMS Chief Strategy Officer Jeff Cohen this session will equip INFORMS members with the tools, tactics, and confidence to engage effectively in public policy conversations that affect research funding, workforce development, AI, healthcare, and more. Using the new INFORMS Advocacy Toolkit as a foundation, this interactive training will walk participants through how to craft compelling messages, build relationships with policymakers, leverage media, and activate advocacy at the local, national, and institutional levels. Whether you’re advocating for your field, your work, or the broader operations research and analytics professions, this session will help turn your technical credibility into public impact.

2 - Advocacy Governance Committee Award

Zachary Collier, Radford University, Salem, VA, United States

Join us as we announce the winner of the 2025 Advocacy Governance Committee Award for Public Engagement The Advocacy Governance Committee (AGC) award for Public Engagement recognizes excellence of an INFORMS member who has actively engaged with the public to advance the profession in the topic areas covered by INFORMS, either by communicating relevant research results or by advocating for use of management science, operations research, and analytics to inform policy decision making. This award is meant to highlight, recognize, and honor the work that INFORMS members do to connect their work to public and private sector decision making processes that can have an impact for society, including multiple stakeholders, in any public policy area.

TC04

Building A Level 3 A304

Financial Strategy, Derivatives & Innovation

Contributed Session

Contributed

Chair: Ali Akbar, University of Science and Technology of China, Department Statistics and Finance, Hefei, N/A

1 - The Predictive Power of The Inventory-to-Sales Ratio

Yifei Wang, New York University, Leonard N. Stern School of Business, New York, NY, United States, Zhengyuan Zhou, Evan Xiao

This study explores the predictive power of the inventory-to-sales ratio (ISR) in forecasting gross margin performance for companies within the S&P 500. By analyzing historical financial data, we examine how changes in ISR impact the profitability and efficiency of firms, particularly focusing on gross margin as a key indicator of financial health. Our findings reveal that a higher or lower ISR can signal potential shifts in cost structures, supply chain effectiveness, and pricing strategies, all of which directly influence gross margin outcomes. The study demonstrates that the inventory-to-sales ratio provides valuable foresight for investors and managers, as it reflects both operational efficiency and market demand dynamics. These results offer a novel approach to financial forecasting, highlighting the ISR as a critical predictor for gross margin trends within large-cap companies in the S&P 500.

2 - An Potential Game Framework for N-Player Dynamic Games

Xinyu Li, UC Berkeley, Berkeley, CA, United States, Xin Guo, Yufei Zhang

This paper proposes and studies a general form of dynamic N-player non-cooperative games called α -potential games, where the change of a player’s objective function upon her unilateral deviation from her strategy is equal to the change of an α -potential function up to an error α . Analogous to the static potential game (which corresponds to $\alpha = 0$), the α -potential game framework is shown to reduce the challenging task of finding α -Nash equilibria for a dynamic game to minimizing the α -potential function. Moreover, an analytical characterization of α -potential functions is established, with α represented in terms of the magnitude of the asymmetry of objective functions’ second-order derivatives. For stochastic differential games in which the state dynamic is a controlled diffusion, α is characterized in terms of the number of players, the choice of admissible strategies, and the intensity of interactions and the level of heterogeneity among players. Two classes of stochastic differential games, namely distributed games and games with mean field interactions, are analyzed to highlight the dependence of α on general game characteristics that are beyond the mean-field paradigm, which focuses on the limit of N with homogeneous players. To analyze the α -NE, the associated optimization problem is embedded into a conditional McKean–Vlasov control problem. A verification theorem is established to construct α -NE based on solutions to an infinite-dimensional Hamilton–Jacobi–Bellman equation, which is reduced to a system of ordinary differential equations for linear-quadratic games.

3 - Investment, Financing, Strategic Debt Service, and Liquidation

Takashi Shibata, Hosei University, Tokyo, Japan, Michi Nishihara, Yuan Tian

We develop a dynamic real options model to examine how the introduction of both collateral and debt renegotiation (i.e., strategic debt service) influences a firm’s investment and financing (capital structure) decisions. Our findings show that the inclusion of these features accelerates investment and reduces debt issuance, resulting in

lower credit spreads and higher leverage. When firms can issue one of two types of collateralized debt—bank (renegotiable) debt or market (non-renegotiable) debt—those with substantial collateral are more likely to prefer market debt over bank debt. These results are consistent with existing empirical evidence.

4 - Vector Dependence and Global Systemic Financial Risk

Robert James, NGS Super/The University of Sydney, Melbourne, Australia, Jessica Leung, Artem Prokhorov

This paper introduces a new measure of global systemic financial risk by modeling the dependence between vectors of equity returns for systematically important banks listed in the U.S., Europe, and Asia. Our approach uses vector copulas to construct distributions with multivariate margins. The new measure of systemic risk captures the degree to which the stock return losses of systematically important banks and insurers are co-monotonic between geographical regions, identifying periods where the global economy is highly susceptible to a financial shock. Crucially, the new measure is robust to changes in the within region dependence. We study the evolution of this new measure around key historical financial stress events and compare it to other dependence based measures of systemic risk from prior literature. Quantile regressions demonstrate that our new measure predicts statistically significant declines in the left tail of the distribution of global financial sector stock returns and real macroeconomic variables, such as global industrial production.

5 - Informed Option Trading: Evidence from Cybersecurity Event Disclosures

Xiaoyu Zhao, University of Massachusetts Amherst, Amherst, MA, United States

This paper investigates the information content of options trading before data breach disclosures. I find abnormal implied volatility (IV) skew prior to the disclosures, a measure that captures the relative expansiveness of put options, negatively and significantly predicts the stock returns around the disclosures. The finding is robust to a placebo test and propensity score matched sample. Moreover, the predictive power of IV skew is stronger when the option market is more liquid relative to the stock market. It is also stronger when the underlying stock under-performs before the disclosures. Overall, my research results confirms the information content of cybersecurity event disclosures and suggests that there is informed option trading prior to such events.

6 - A Dynamic Model for Open Banking

Ranu Castaneda Medina, University of Alberta, Edmonton, AB, Canada, Christoph Frei

First introduced in the United Kingdom, open banking has been adopted or explored in various countries worldwide. Recent efforts in the USA, including proposed rules by the Consumer Financial Protection Bureau, signal a growing commitment to fostering competition and data portability in financial services. Open banking allows customers to share their financial data with third-party providers, such as fintech companies, thereby increasing competition and expanding access to financial services. While this shift promises benefits such as improved product offerings and expanded financial inclusion, it also introduces significant technical, social, and economic risks. We develop a continuous-time model using a search-and-match framework to analyze borrower heterogeneity and competition between banks and fintechs. Our preliminary analysis shows that open banking reshapes financial services by leveling the playing field between banks and fintechs. However, while it enhances competition, it could also over-empower fintechs, potentially reversing the benefits to borrowers.

TC05

Building A Level 3 A305

Service Science Award Session II

Award Session

Service Science

Co-Chair: Renyu Zhang, The Chinese University of Hong Kong, 12 Chak Cheung Street, Hong Kong, N/A

Co-Chair: Weiwei Chen, Rutgers University, Piscataway, NJ, 08854, United States

1 - Glass Half Full: Strategic Packaging Decisions and Environmental Outcomes

Neda Mirzaeian, University of Texas at Dallas, Richardson, TX, United States, Yanhan (Savannah) Tang

Amid rising environmental concerns and regulatory pressures, firms, particularly beverage manufacturers, face complex trade-offs when choosing among packaging options such as non-recycled plastic, recycled plastic, and glass. While glass is often perceived as the most sustainable option, its true environmental impact depends critically on return and reuse rates, which are influenced by firm-led deposit-refund systems and sustainability advertising. We develop an analytical model to study a firm's joint decisions on packaging type, pricing, and advertising efforts in a setting with heterogeneous consumer environmental awareness and greenwashing risk. We quantify the conditions under which glass can outperform plastic in terms of profit and emissions—or, paradoxically, lead to higher environmental harm. Under optimal pricing and advertising efforts, the firm can achieve a rare “win-win-win”: prevent potential greenwashing while simultaneously improving profits and customer utility. We extend the model to examine the effects of recycled content mandates and endogenous deposit values, showing how firms can balance deposit incentives and advertising strategies to engage both environmentally aware and unaware consumers. A calibrated numerical study using data from a regional U.S. dairy firm reveals that adopting the model's recommended policy—glass packaging with a premium price and targeted advertising—can increase profits by almost 60% while reducing emissions by over 30%. Our findings can help identify when glass containers outperform plastic alternatives and guide how sustainability advertising and refundable deposits can be leveraged to boost glass return rates, improve environmental outcomes, enhance customer utility, and increase profits.

2 - Subsidizing Social Welfare Programs: Contracted Slots or Vouchers?

Priyank Arora, University of South Carolina, Columbia, SC, United States, Senay Solak

Subsidy welfare programs rely on the private service providers' participation to offer services to beneficiaries and, in turn, generate societal outcomes. The voucher program is a widely used program wherein providers serve beneficiaries using their remaining capacity after serving private customers. However, overall service quality for beneficiaries may be low due to limited restrictions on participating providers' quality, and reimbursement payments for providers are uncertain as beneficiaries often do not use the offered vouchers. Alternatively, the contracted slot program restricts participation to high-quality providers who must reserve some capacity to serve beneficiaries (for guaranteed reimbursements). In each program, since providers' benefit of mitigating demand uncertainty by serving beneficiaries differs, providers' decisions and equilibrium outcomes could differ significantly. We develop an analytical model to characterize the optimal participation strategy of providers in each program and identify conditions when one program outperforms the other from the beneficiaries' and providers' perspectives. Our findings on the optimal participation strategies of providers imply that: Providers should carefully reevaluate their decisions if the government switches program types. More importantly, our comparative analysis of societal outcomes for beneficiaries and total providers' profit under these programs imply that: While practitioner debates on these programs typically focus on the number of providers, our findings underscore the importance of considering other localized factors when evaluating program outcomes. Further, based on Massachusetts' child care subsidy programs, our case study illustrates that the contracted slot (versus voucher) program can generate approximately 16% greater societal outcomes and 4% higher total providers' profit.

3 - The Impact of Recommender Systems on Content Consumption and Production: Evidence from Field Experiments and Structural Modeling

Zhiyu Zeng, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Zhiqi Zhang, Tat Chan, Dennis Zhang

Online content-sharing platforms such as Facebook and TikTok, where users act as both content creators and consumers, have become integral to daily life, relying on complex algorithms to recommend user-generated content (UGC). While prior research and industry efforts have primarily focused on designing recommender systems to encourage users to consume more content, the impact of the content recommended to users on their own content production remains understudied. To address this gap, we conducted a large-scale randomized field experiment on a major video-sharing platform. Users were randomly assigned to either a treatment or control condition, and we manipulated content recommendations by excluding a subset of highly popular creators from being recommended to treated users. This reduced recommendation quality of treated users compared to control counterparts. Our results indicate that this decline led to a 1.34% decrease in video-watching time but a 2.71% increase in videos uploaded by treated users. These findings reveal a trade-off in recommender system design: while higher recommendation quality boosts consumption, it may suppress production. To optimize recommendations, we developed a structural model in which users' choices between consuming and producing content are inversely affected by recommendation quality. Counterfactual analyses suggest that slightly lowering recommendation quality—rather than prioritizing the most popular creators—can maximize the overall value of both production and consumption for platforms. This highlights the need to balance content consumption and production when designing

recommender systems on these social media platforms. Keywords: User-Generated Content, Recommender System, Field Experiment, Structural Model, Platform Operations, Production

4 - Information Provision In Platform Competition

Wenchang Zhang, Indiana University Bloomington, Bloomington, IN, United States, Kostas Bimpikis, Yiangos Papanastasiou

This paper studies how platform competition shapes information provision strategies in freelance marketplaces. We introduce a novel steady-state model in which competing platforms choose how to disclose freelancer quality information to heterogeneous consumers. Although the problem of computing competitive equilibria that involve endogenous information provision is inherently infinite-dimensional and seemingly intractable, we demonstrate that it can be reduced to a finite-dimensional convex quadratic program. This structure allows us to solve for competitive equilibria efficiently, extracting the main effects of platform competition. When the number of competing platforms is small, we find that equilibrium information policies involve significant levels of obfuscation in order to create the most profitable “product lines” (as in the case of platform monopolies). However, as the number of competing platforms increases, we show that platforms adopt increasingly transparent information policies. This transparency improves the efficiency of matching between service providers and consumers and decreases transaction prices, increasing total welfare and consumer surplus, but significantly decreasing total platform revenues. Furthermore, in markets characterized by asymmetric platform costs, specialization emerges, with lower-cost (higher-cost) platforms concentrating on higher-quality (lower-quality) freelancers. At the same time, asymmetric decreases in platform costs do not necessarily translate to increases in total welfare or consumer surplus.

TC06

Building A Level 3 A306

Analytical Modeling in eBusiness

Invited Session

eBusiness

Chair: Yeongin Kim, Wake Forest University, Glen Allen, VA, United States

1 - Strategic Rating Effort and Platform Revenue in Gig Markets: A Stackelberg Mean Field Game Approach

Joohyun Kim, University of North Carolina at Charlotte, Charlotte, NC, United States, Monica Johar, Vijay Mookerjee

Digital labor platforms have redefined traditional employment by introducing decentralized, task-based work arrangements, where platform ratings play a central role in shaping worker incentives and outcomes. In this paper, we model a dynamic competition among gig workers who strategically choose their continuous effort levels to improve platform ratings that depend on both their own efforts and those of other workers. Meanwhile, the platform sets its commission rate based on the aggregate rating outcome, aiming to optimize revenue. To analyze this interdependent system, we formulate a Stackelberg Mean Field Game (MFG) that captures the strategic interactions between a large population of homogeneous agents and the platform. Within a linear-quadratic control setting, we derive closed-form expressions for agents' optimal effort levels and their steady-state rating trajectories. We also determine the platform's optimal commission rate under steady-state dynamics. Finally, we quantify the inefficiency arising from decentralized decision-making through the Price of Anarchy (PoA), highlighting the gap between individual incentives and socially optimal outcomes. Our findings offer insights into three key questions: (1) how agents adjust their effort in response to rating-based incentives when dynamically interconnected; (2) how platform-designed commission policies influence long-term agent engagement and revenue; and (3) how PoA can guide platform design to mitigate inefficiencies in gig ecosystems.

2 - Phantom Capacity: Informational Cost of Multihoming on Ride-Hailing Platforms

Jaelynn Oh, Wake Forest University, Winston-Salem, NC, United States

This paper examines the phenomenon of multihoming among drivers on ride-hailing platforms—where drivers simultaneously operate on multiple platforms. While multihoming can obscure real-time availability, leading to inefficiencies in matching supply with demand, it also enhances overall system efficiency by effectively pooling driver capacity across platforms. We investigate this fundamental trade-off between reduced matching accuracy and increased resource availability.

3 - A Game-Theoretic Analysis of Pricing in Blockchain-Based Storage Sharing Platforms

Kun Soo Park, Seoul National University, Seoul, Korea, Republic of, Youngjae Jang, Jaegung Sim, Daegon Cho

As data storage demands grow, blockchain-based peer-to-peer (P2P) storage platforms offer alternatives to traditional cloud services but face challenges in ensuring reliable availability. This study analyzes how three pricing schemes—two-part tariff, subscription, and hybrid—affect short-term profits and overall welfare in P2P markets. Using a game-theoretic model, we find that two-part tariffs maximize platform profits, while subscription pricing benefits renters most. Total welfare depends on provider availability: subscription excels when providers are plentiful, whereas hybrid or two-part tariffs perform better with scarcity. These findings guide early-stage platforms in balancing profit and user growth based on provider and renter dynamics.

4 - When Data Can't Flow Upstream: Implications of API Training Restrictions on GenAI Providers

Zhe Wang, City University of Hong Kong, Hong Kong, Hong Kong, Yi Gao, Xinxin Li

This study examines the strategic and welfare implications of restricting generative AI (GenAI) companies from using API-generated data for model training. Motivated by emerging debates and the conflicting reality, we develop a game-theoretic model featuring a foundation model provider, a developer, and consumers with heterogeneous preferences. The provider offers its base model through API licensing to the developer, who builds upon it to provide a differentiated service. We compare two regimes: a regime where the provider can use API-generated data for training, and a restricted regime where such usage is prohibited. Our analysis reveals that restricting API data for training may paradoxically enhance the foundation model's service quality. However, we also show that under certain conditions, despite alleviating privacy concerns, such a restriction may reduce consumer surplus by limiting product differentiation. Our findings offer actionable insights for AI providers, developers, and policymakers navigating data governance in GenAI ecosystems.

TC07

Building A Level 3 A307

Complex Systems: Energy, Mobility, and Urban Intelligence

Contributed Session

Contributed

Chair: Jacob Zijian Zeng, Gonzaga University, 1605 N River Ridge Blvd, Spokane, WA, 99224, United States

1 - A Physically and Socially Inspired Hybrid Model for Trajectory Prediction of Neighboring Vehicles in Autonomous Driving

Yuhao Li, School of Transportation, Southeast University, Nanjing, China, People's Republic of, Wang Jian, Zhening Liu

Predicting the motions of surrounding human-driven vehicles is vital for path planning and collision avoidance of autonomous driving systems in mixed flow scenarios. However, accurately predicting the trajectories of surrounding vehicles still remains as a challenging research issue due to the complex dynamics of vehicle interactions and heterogeneity of different drivers' driving styles in real-world driving scenarios, particularly when striving for socially and compliant personified predictions. To tackle this challenge, we propose an innovative multimodal hybrid model for collaboratively predicting the trajectories of neighboring vehicles. Inspired by the Potts model in physics, we perform graph-structured encoding on the directly obtainable features from each neighboring vehicle, including position, velocity, acceleration, as well as drivers' personalized characteristics such as driving style and intention. Moreover, the concept of Blau Space from sociology is incorporated to aggregate the encoding results of these heterogeneous features. The aggregated features are then processed through several Spatio-Temporal Graph Convolutional Network (STGCN) layers and spatio-temporal transformer layers to capture the dynamic interactions among neighboring vehicles and across the global traffic environment, respectively. Finally, the candidate predicted trajectory with the highest probability is selected based on the negative logarithmic likelihood calculation results. Experimental results on the NGSIM

and highD datasets demonstrate that the proposed model outperforms seven selected baseline models. Furthermore, ablation studies are conducted to validate the effectiveness of model design.

2 - A Real-TIME Control to Connected Automated Vehicles Operation in Roundabouts

Abdullah Farabi, NC State University, Raleigh, NC, United States, Ali Hajbabaie

This study presents an efficient control framework for connected and automated vehicles (CAVs) operating in roundabouts. A throughput-maximizing optimization model is developed that utilizes vehicle state information (position and speed) to determine optimal safe trajectories. The model, being non-linear, requires efficient techniques for practical implementation. First, the problem is simplified using Frenet coordinates, reducing the two-dimensional spatiotemporal domain to a single dimension. A safe trajectory construction heuristic is then applied to solve the simplified model. This heuristic ranks CAVs based on their proximity to potential conflict points within the roundabout and generates trajectories sequentially according to this priority. The constructed trajectories satisfy both kinematic feasibility and safety constraints. To address uncertainties in the system, a receding horizon control framework is integrated into the approach. The proposed framework is evaluated against conventional roundabout operations. A case study results show improvements across several performance metrics, including delay, number of stops, travel time, and throughput. Moreover, the framework, exhibiting computational efficiency, supports its potential use for real-time deployment. Overall, this research contributes a practical and scalable solution for enhancing roundabout operation through an intelligent CAV coordination framework.

3 - Accelerating the Net-Zero Pathway via Technology-Driven Transition: A CGE-Based Analysis of the Steel Sector

Yoonha Lee, Seoul National University, Seoul, Korea, Republic of, Yoonmo Koo

This study applies a dynamic Computable General Equilibrium (CGE) model to analyze Korea's steel sector transition under carbon neutrality targets. The model structurally differentiates three production technologies—BF-BOF, DRI-EAF, and H₂-DRI-EAF—by assigning each as an individual sector and nesting them within a CES production structure. Although technological change is not endogenously determined in standard CGE frameworks, the model design introduces a competition-like mechanism by embedding distinct input characteristics based on disaggregated data. This setup allows the simulation of technology diffusion without relying on exogenously imposed adoption rates, distinguishing it from conventional approaches.

The scenario analysis includes a baseline reflecting Korea's 2050 net-zero power mix, along with policy packages that combine carbon pricing and technology-specific efficiency improvements.

Results indicate that carbon pricing alone induces only limited technological shifts. However, when combined with efficiency gains in hydrogen production and steelmaking, the hydrogen-based route achieves a meaningful market share by 2050 without imposing significant economic costs. Emissions are substantially reduced while macroeconomic performance remains stable, underscoring the importance of coordinated policy mixes over standalone instruments.

4 - Navigating the Future of Transportation Technologies: Translating Expert Knowledge into Actionable Scenarios

Farah Naz, University of Texas at Arlington, Arlington, TX, United States, Stephen Mattingly, Kate Hyun

The rapid evolution of transportation technologies necessitates a proactive approach from agencies to navigate an uncertain future. Given the scarcity of historical data for predictive modeling, understanding subject matter experts' (SMEs) perspectives is crucial for aligning infrastructure and policy with societal and economic objectives. This study introduces a framework to translate SME insights, gathered via surveys, into practical scenarios that guide future transportation planning.

The research employs hierarchical clustering, an unsupervised machine learning technique, to develop unbiased scenarios. Initially, SME responses were grouped based on their answers across four key attributes: preferred technologies, anticipated goals, effective solution strategies, and impactful technologies. A subsequent granular analysis examined each attribute individually to identify any overlooked scenarios. The final scenarios are derived from both the complete dataset and its subsets, emphasizing significant patterns within crucial clusters.

To manage mixed data types, the Gower distance method was used, handling both categorical and numerical data effectively. The Ward's minimum variance method (Ward2) served as the linkage function, minimizing within-cluster variance and resulting in six distinct clusters. Validation techniques including Average Silhouette Width, Gap Statistics, and the Calinski-Harabasz Index confirmed the optimal number of clusters and consistency of the derived scenarios, each illustrated with practical examples.

These scenarios offer a structured framework for stakeholders in transportation planning and policymaking, aligning technology adoption with expert insights and strategic goals. This study offers a methodological contribution to scenario analysis and a thorough understanding of expert-driven forecasts in transportation technology planning, providing strategic insights for future development.

5 - Efficiency Gains OR Behavioral Losses? Experimental Evidence of the Energy Rebound Effect Among Polish Participants

Umar Ashraf Rana, Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland, Bartlomiej Dessoulavy-Sliwinski

This study experimentally investigates the rebound effect arising from adopting energy-efficient technologies, introducing a novel production-oriented framework primarily focused on Poland, supplemented by comparative data from Spain, Uruguay, and the Dominican Republic. Departing from conventional consumption-focused approaches, we employ a laboratory-based Two-Factor Production Technology task (2FPT). Participants engage in solving visual puzzles under constrained time and brightness conditions, with brightness levels serving as a proxy for costly energy consumption decisions.

The experiment comprises ten rounds divided into baseline (fixed brightness price) and two treatment conditions: a reduced price scenario (PD) and a reduced price scenario incorporating explicit environmental externalities (PD-E). Specifically, we examine whether Polish participants under the PD condition significantly increase brightness consumption without corresponding gains in productivity (H1), and whether integrating environmental externalities into pricing (PD-E) effectively curbs this rebound behavior (H2). Additionally, we assess cross-country heterogeneity (H3), hypothesizing greater responsiveness to externalities in countries with higher renewable energy usage. Finally, the study explores cognitive factors, predicting a negative correlation between inefficient brightness consumption and participants' energy literacy and fluid intelligence scores (H4).

Focusing primarily on Poland, this research sheds light on behavioral mechanisms underlying energy efficiency adoption, the effectiveness of environmental externalities in mitigating rebound effects, and the role of individual cognitive characteristics in shaping sustainable production decisions.

6 - Algorithmic Nudging for Energy Savings and Environment: Evidence from An Iot Platform

Jacob Zijian Zeng, Gonzaga University, Spokane, WA, United States, Serasu Duran, Nil Karacaoglu, Nur Sunar

We study how algorithmic nudging can reduce energy consumption using data from an IoTplatform.

TC08

Building A Level 3 A310

Smart Sustainable Mobility

Invited Session

Information Systems

Chair: Julius Kuhmann, University of Cologne, Düsseldorf, N/A, Germany

1 - Cnn-Based Path Planning for Unmanned Aerial Vehicle Routing

Matthias Winkenbach, Massachusetts Institute of Technology, Cambridge, MA, United States, Rodrigo Hermosilla

Unmanned aerial vehicles (UAVs) for last-mile delivery face complex routing challenges in urban environments, including regulatory altitude restrictions, obstacle avoidance, and noise mitigation requirements. We present a novel application of convolutional neural networks (CNNs) with spatial attention mechanisms to optimize UAV routing decisions while maintaining compliance with urban airspace regulations. Our approach extends existing CNN-based path planning by incorporating altitude awareness into the path finding process, transforming the traditional 2D planning problem into a more nuanced 3D routing framework. The model accounts for various regulatory constraints derived from aviation guidelines, including altitude limitations, clearance requirements around structures, and designated no-fly zones. The proposed architecture employs a U-Net backbone with spatial attention mechanisms that dynamically highlight critical routing information while maintaining awareness of the three-dimensional environment. By training on high-quality solutions derived from traditional optimization methods, our CNN can generate compliant flight paths with inference times suitable for real-time applications. This research bridges the gap between theoretical path planning algorithms and practical urban airspace management, offering a scalable solution for future UAV delivery operations. The approach balances computational efficiency with regulatory compliance, providing a foundation for safe, efficient autonomous aerial delivery in complex urban environments.

2 - Joint Inventory Routing for EV Battery Swapping Networks with Hybrid Charging

Long He, George Washington University, Washington, DC, United States, Ningwei Zhang, Shanshan Wang, Yuli Zhang

We study a joint inventory-routing problem for electric vehicle (EV) battery swapping systems that feature hybrid charging, where batteries may be charged either on-site at local stations or centrally at a depot, under uncertain vehicle arrival rates and stochastic battery charging durations. The system comprises a central charging station and multiple battery swapping stations, each experiencing random demand and constrained by service-level requirements. Returned batteries with short charging times are charged locally, while those exceeding a threshold are routed to the central station, introducing a two-echelon logistics structure with state-dependent flows.

We develop a two-stage optimization model that simultaneously determines the battery delivery routes and inventory replenishment levels across the network. The model incorporates a probabilistic stockout constraint to ensure high service quality and accounts for the impact of hybrid charging on inventory pooling and routing efficiency. To address the problem's non-convexity and computational complexity, we derive closed-form approximations for service-level constraints and reformulate the problem using second-order cone programming. We further propose an efficient branch-and-cut algorithm enhanced with extended polymatroid cuts to improve tractability and scalability.

Our model generalizes existing battery logistics frameworks by integrating hybrid service modes, stochastic charging durations, and routing constraints into a unified optimization framework. Numerical results demonstrate the operational benefits of hybrid charging and validate the effectiveness of the proposed solution approach. This work provides a tractable, analytically grounded method for managing EV battery logistics in urban mobility systems under uncertainty.

3 - Measuring the Value of Information Systems for Virtual Power Plants: the Case of Electric Vehicle Aggregators

Diego Cebreros, CentraleSupélec Université Paris Saclay, Gif-sur-Yvette, France, Jan Lepoutre, Yannick Perez

The transition toward low-carbon energy systems increasingly relies on digital solutions enabling flexible use of distributed energy resources. One promising avenue is aggregating Electric Vehicle (EV) batteries into Virtual Power Plants (VPPs), allowing coordinated charging and discharging to support grid stability and renewable energy integration. This paper investigates the role of information systems (IS) in enabling such coordination, focusing on how enhanced communication affects system performance and value creation. We develop a simulation framework modeling interactions among EV aggregators, electricity system operators, and EV users. The framework captures operational constraints—such as grid reliability and market signals—and user mobility requirements. By evaluating scenarios beyond current communication standards, we quantify the value of different scenarios of information exchanges. Our findings show that improved IS can significantly increase aggregator profitability—doubling it in some cases—while better aligning energy dispatch with renewable generation. This study contributes theoretically by introducing a novel simulation-based framework to assess the value of information in settings where EV owners' mobility behavior is independent across actors, rendering statistical knowledge of the fleet insufficient for accurate prediction. It also has practical implications for IS architecture design in energy-mobility integration and provides insights for policymakers and firms developing IS standards for VPPs.

4 - Estimating the Effects of Congestion Pricing Using Predicted Counterfactuals

Julius Kuhmann, University of Cologne, Cologne, Germany, Gordon Burtch, Wolfgang Ketter

Evaluating the effects of policy interventions in urban mobility can present methodological challenges, particularly due to the lack of suitable control observations. Even when interventions are geographically localized, the highly interconnected nature of mobility systems often precludes surrounding areas from serving as valid counterfactuals. Consequently, many conventional methods from the causal inference toolkit cannot be readily applied in such settings.

We address this issue in the context of New York City's congestion pricing program, implemented in early 2025, which introduced a fee for vehicles entering Manhattan's central business district. To estimate the program's short-term impact on mobility behavior, we adopt a machine learning-based approach that predicts counterfactual outcomes using historical pre-intervention data. This method allows us to gain insights on the effect of the policy despite the absence of a traditional control group.

Focusing on ride-hailing activity, we document substantial changes in ridership volumes following the policy's implementation. Our findings demonstrate the potential of predictive, data-driven methods to support causal inference in complex urban environments where traditional approaches fall short.

TC09

Building A Level 3 A311

Machine Learning and GenAI

Invited Session

Information Systems

Chair: Sorouralsadat Fatemi, California State University Monterey Bay, Seaside, CA, United States

1 - From Lexicons to Generative AI: Benchmarking Data Annotation in Business Research

Reza Mousavi, University of Virginia, Charlottesville, VA, United States

Reza Mousavi (panelist)

2 - Should You use LLMs to Simulate Opinions? Quality Checks for Early-Stage Deliberation

Terrence Neumann, University of Texas at Austin, Austin, TX, United States

The emergent capabilities of large language models (LLMs) have sparked interest in assessing their ability to simulate human opinions in a variety of contexts, potentially serving as surrogates for human subjects in opinion surveys. However, previous evaluations of this capability have depended heavily on costly, domain-specific human survey data, and mixed empirical results about LLM effectiveness create uncertainty for managers about whether investing in this technology is justified in early-stage research. To address these challenges, we introduce a series of quality checks to support early-stage deliberation about the viability of using LLMs for simulating human opinions. These checks, designed to evaluate Likert-scale opinion data, emphasize **logical consistency**, **prompt stability**, and **alignment with stakeholder expectations of model outputs**, thereby reducing dependence on human-generated data in the initial stages of evaluation. We demonstrate the usefulness of the proposed quality control

tests in the context of AI-assisted content moderation, an application that both advocates and critics of LLMs' capabilities to simulate human opinion see as a desirable potential use case. None of the tested models passed all quality control checks, revealing several failure modes. We conclude by discussing implications of these failure modes and recommend how organizations can utilize our proposed tests in their risk management practices when considering the use of LLMs for opinion simulation. We make our crowdsourced dataset of claims with human and LLM annotations (**TopicMisinfo**) publicly available for future research.

3 - AI Generated Content and Home Bias: Answer Inclination on Q&A Websites

Vincent Zhao, University of Florida, Gainesville, FL, United States, Jingchuan Pu

Knowledge Sharing or Q&A websites are popular venues for users to seek professional help in answering their questions. Answering questions is a non-trivial task that is significantly costly in terms of time and cognitive effort. With the introduction of AI Generated Content (AIGC) from large language models (LLMs) such as ChatGPT, the cost for providing answers has dramatically decreased. Users can use LLMs to help research, write, and format responses reducing cognitive burden. In this paper, we identify the effect that the presence of LLMs has on user home bias in a Q&A website. We show that after the existence of ChatGPT, users' home bias in answering questions has significantly changed.

TC10

Building A Level 3 A312

Digital Innovation and Technology

Contributed Session

Contributed

Chair: Jing Hou, California State University, San Marcos, San Marcos, CA, United States

1 - How Does a Successful Initial Coin Offering (Ico) Event Affect Blockchain-Related Firm's Stock Market Return?

Lingyu Li, Virginia Commonwealth University, Richmond, VA, United States

This paper studies the impact of successful ICO events on the equity returns of blockchain-related firms. We use 494 successful ICOs from 2018 to 2023 and estimate cumulative abnormal returns (CARs) over a [-30, +30] days event window using the market model, Fama-French 3-factor, and 5-factor models. Results show significant positive CARs, with 0.227% on the event day and 0.495% over the 30-day window (Fama-French 3-factor), indicating that the blockchain stock market positively values ICOs. Random effects regression reveals stronger positive reactions when ICOs exhibit greater rating variance and raise more capital, while bonus offerings in the ICOs are associated with weaker reactions. These findings suggest that perceived evaluation credibility and financial commitment enhance market valuation, whereas offering bonus in ICOs may signal lower project quality and reduce market valuation.

2 - The Differential Effect of Developer Expertise on The Knowledge Creation in Open Source Software (OSS) Development

Orunc Temizkan, Özyeğin University, Istanbul, Turkey, Ram Kumar

OSS projects depend on contributions from software developers with expertise in different knowledge areas to facilitate knowledge creation. Prior research has illustrated that network structure is related to knowledge creation in the context of OSS development. However, the effect of developer expertise is under researched. A developer may have deep knowledge in a specialized knowledge area as well as broad knowledge in several generalized knowledge areas. Knowledge transfer between developers to facilitate project success depends on the depth and breadth of knowledge that individual developers have as well as their position in the network. We examine the effect of developer expertise through their contribution to different activities on project success in the context of OSS development. We also empirically illustrate the differential impact of developer expertise (in terms of breadth and depth of knowledge) on project success. Data collected from GitHub projects, which is a leading OSS hosting website, is used to illustrate these effects. Research and managerial implications will also be discussed.

3 - A Data-Centric Framework for Reproducible Machine Learning in Interdisciplinary Collaborations

Zhiwei Li, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States, Carl Kesselman, Tran Huy Nguyen, Benjamin Y. Xu, Kyle Bolo

As machine learning (ML) becomes increasingly central to high-impact domains such as healthcare, concerns about reproducibility and social consequences have grown. ML development often involves interdisciplinary teams of domain experts, ML engineers, and data managers who iterate over data, models, and decisions. This exploratory process introduces challenges in maintaining consistency, traceability, and accountability, especially when early-stage errors propagate across the lifecycle and undermine confidence in outcomes.

This work introduces a data-centric framework that supports reproducibility by treating data artifacts as evolving rather than static. The framework is centered on six core artifact types—Dataset, Feature, Workflow, Execution, Asset, and Controlled Vocabulary—which are independently versioned and tightly linked to unify activities across the ML lifecycle. By modeling the relationships among these artifacts, the framework captures the evolution of data, code, and decisions across roles, enabling teams to trace modeling rationale and revisit past iterations. This structure aligns human and technical workflows through traceability, transparency, and decision accountability.

The framework is implemented in Deriva-ML, a data-centric information system that automates full-lifecycle experiment tracking. We evaluated our approach in a clinical application that uses AI models and retinal images to diagnose glaucoma, supporting tasks such as feature extraction, manual label refinement, and robustness evaluation. This evaluation demonstrates how the framework can enhance reproducibility and interdisciplinary collaboration in ML workflows across domains where reliable, transparent decision-making is critical.

4 - Cohesion OR Equivalence? the Social Network Mechanisms Driving Innovation Diffusion Among Talents in Quantum Communication

Zhuyan Hui, Renmin University of China, Beijing, China, People's Republic of, Zhiyang Liu, Xuansheng Wang, Zengxiang Shi, Lifan Chen

Quantum communication, due to its theoretically absolute security, has become a pivotal area in international competition for safeguarding national information security. Attracting and continuously stimulating innovative talents is crucial for advancing this field, yet the underlying mechanisms remain underexplored. Grounded in social influence theory, this study examines how social influence within research networks affects talents' entry into the quantum communication field and their innovative productivity, and further investigates the moderating role of triple helix (government-industry-university) collaboration. Analyzing panel data of 1,160 researchers across ten countries from 2016 to 2021, we reveal three primary findings: (1) Innovations produced by collaborators and structurally equivalent peers positively influence focal researchers' own productivity; (2) Strong government-industry-university collaboration negatively moderates the influence of collaborators' innovation outputs on focal researchers' productivity, but has limited effect on structurally equivalent peers; (3) Both collaborator and structurally equivalent peer innovations significantly encourage new talent entry into quantum communication, with stronger triple helix collaboration enhancing this effect. This research contributes to understanding the dynamics of talent acquisition and innovation diffusion in quantum communication, providing practical guidance for institutional collaboration. Specifically, we uncover a dynamic "early-stage stimulation and later-stage inhibition" pattern in the role of government-industry-university cooperation on talent development, suggesting the need for dynamically adjusting collaboration strategies in line with national development stages in the quantum communication field to optimize innovation outcomes.

5 - Tipping vs. Subscription System on Video Streaming Platforms: a Game-Theoretic Analysis

Tomoyuki Ichikawa, Keio University, Yokohama, Japan, Nobuo Matsubayashi

In recent years, many video streaming platforms have introduced tipping systems, such as YouTube's Super Chat and TikTok's Gift. These systems allow consumers to donate to content creators, and the platforms earn revenue by receiving a portion of the tips as a commission. This study conducts a game-theoretic analysis to investigate the profitability of the tipping system for platforms and content creators compared to the subscription system. Specifically, we analyze an equilibrium of game where two symmetric platforms can adopt either a tipping or subscription system. Our results reveal that platforms are likely to benefit from adopting tipping systems under a

duopoly and subscription systems under a monopoly. Additionally, we show that, under a duopoly and when content is less differentiated and the number of content creators is large, the tipping system achieves a win-win outcome for platforms and content creators.

6 - Don't Kill the Game: Strategic Cheat Pricing Under Network Effects

Jing Hou, California State University, San Marcos, San Marcos, CA, United States, Thanh Nguyen

The rapid growth of online gaming has been accompanied by a significant rise in cheating, driven by a market for third-party tools that give users unfair advantages through cheating add-on software. Unlike traditional malicious hackers who seek to disrupt target systems, cheat developers operate parasitically: their profits depend on the continued participation of honest players and the long-term viability of the game ecosystem. Excessive cheating degrades gameplay quality, prompting regular players to leave and ultimately reducing the profitability of cheating software.

This study develops a game-theoretic model to examine the optimal pricing strategy for cheat developers in the presence of both cheating and non-cheating players. The model incorporates three network effects: (1) a positive direct effect, where gameplay becomes more enjoyable as the number of regular players increases; (2) a negative indirect effect, where the prevalence of cheating reduces enjoyment for regular players; and (3) a positive indirect effect, where the value of cheating increases with the number of regular players. We model a Stackelberg game in which the cheat developer acts as a leader, setting prices while anticipating the strategic responses of heterogeneous players. Our findings offer new insights into the dynamics of cheating in online games and illustrate how cheat developers are strategically exploiting the gaming ecosystems to maximize their profit.

TC11

Building A Level 3 A313

TSL Stella Dafermos Mid-Career Award Lecture, and Remembering Past TSL Herman Lifetime Achievement Award Winner Hani Mahmassani

Award Session

Transportation Science and Logistics (TSL)

Chair: Karen Smilowitz, Northwestern University, Evanston, IL, United States

Co-Chair: Yue Wang, The University of Arizona, Tucson, AZ, United States

1 - TSL Stella Dafermos Mid-Career Award Lecture

Karen Smilowitz, Northwestern University, Evanston, IL, United States

TSL Stella Dafermos Mid-Career Award Lecture

2 - Remembering Past TSL Herman Lifetime Achievement Award Winner Hani Mahmassani

Karen Smilowitz, Northwestern University, Evanston, IL, United States

This time will be a tribute to Hani Mahmassani with memories from Srini Peeta, Anna Nagurney, Mike Hyland, Marco Nie, and Karen Smilowitz

3 - Remembering Past TSL Herman Lifetime Achievement Award Winner Hani Mahmassani

Srinivas Peeta, Georgia Tech, Atlanta, GA, United States

This time will be a tribute to Hani Mahmassani with memories from Srini Peeta, Anna Nagurney, Mike Hyland, Marco Nie, and Karen Smilowitz

4 - Remembering Past TSL Herman Lifetime Achievement Award Winner Hani Mahmassani

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

This time will be a tribute to Hani Mahmassani with memories from Srini Peeta, Anna Nagurney, Mike Hyland, Marco Nie, and Karen Smilowitz

5 - Remembering Past TSL Herman Lifetime Achievement Award Winner Hani Mahmassani

Michael Hyland, UC Irvine, Irvine, CA, United States

This time will be a tribute to Hani Mahmassani with memories from Srini Peeta, Anna Nagurney, Mike Hyland, Marco Nie, and Karen Smilowitz

6 - Remembering Past TSL Herman Lifetime Achievement Award Winner Hani Mahmassani

Marco Nie, Northwestern University, Evanston, IL, United States

This time will be a tribute to Hani Mahmassani with memories from Srini Peeta, Anna Nagurney, Mike Hyland, Marco Nie, and Karen Smilowitz

TC12

Building A Level 3 A314

Autonomous and Connected Vehicle Technologies

Contributed Session

Contributed

Chair: Rahul Bhadani, The University of Alabama in Huntsville, Huntsville, AL, United States

1 - Path Planning for Tethered Unmanned Systems in Partially Known Environments

Sachet Khatiwada, Clemson University, Clemson, SC, United States, Pamela Murray-Tuite, Matthias J. Schmid

The use of unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs) in a tethered system benefits both vehicles. The advantages for the UAV are the possibility to off-load computations to the UGV and siphon power from the UGV to increase flight time. The UGV benefits from the information obtained by the UAV from a higher vantage point, which is beneficial in partially known environments. We propose a path planning framework for such systems operating in partially known environments. We use the a priori information about the environment to generate a collision-free global path for the tethered UAV-UGV system. Then we rapidly replan to avoid newly detected obstacles as the tethered system moves along the original global plan. We use a hybrid approach that benefits from the rapid exploration of the RRT* algorithm to generate the global plan, then we leverage the random tree it generates alongside a mixed integer quadratic programming model to rapidly replan for a collision-free final path.

Distribution Statement A. Approved for public release; distribution is unlimited. OPSEC # 9684

2 - Mixed Traffic Assignment Under Connected and Autonomous Environment: Model and Algorithm

Zhenning Liu, Southeast University, Nanjing, China, People's Republic of, Wang Jian, Yuhao Li

With the increasing penetration of connected and autonomous vehicles (CAVs), traffic systems will involve a mix of human-driven vehicles (HDVs) and CAVs. Due to the significant differences between human-driven vehicles and CAVs in driving characteristics and route choice behaviors, traffic dynamics and network flow will undergo profound changes. This study develops a mixed traffic assignment model that incorporates both HDVs and CAVs, along with a corresponding fast-solving algorithm. The model is built upon an analysis of mixed traffic flow characteristics, focusing on how CAV platooning and aggregation affect road capacity and travel impedance. HDV users and CAV users follow different route choice principles, described by the stochastic user equilibrium model and the user equilibrium model, respectively. To address the low precision and slow convergence issues in traditional mixed traffic assignment methods, a revised gradient projection algorithm is proposed, which incorporates second-order derivative information to enhance convergence speed, improve solution accuracy, and reduce computational time. Numerical examples in the Sioux-Falls

network and Ningbo city networks demonstrate that the proposed revised gradient projection method achieves an order-of-magnitude improvement in accuracy over the route swapping algorithm. Compared to existing algorithms, the computation time for the same level of accuracy is reduced by 91.1% and 92.4%, respectively.

3 - Modeling the Fundamental Diagram in Pure Cav Multi-String Traffic Environment with Lane-Changing Considerations

Ruining Yang, Georgia Institute of Technology, Atlanta, GA, United States, Yongyang Liu, Srinivas Peeta

Recent advances in connected autonomous vehicles (CAVs) provide a promising opportunity for leveraging automated vehicle platoons to improve roadway capacity, driving safety, and energy efficiency. However, existing studies fail to consider the impacts of lane-changing maneuvers on the dynamics of platooning, as such maneuvers tend to disrupt platoons and increase the inter-platoon spacing for speed harmonization across lanes. As a result, the heterogeneous platoon sizes and non-uniform inter-platoon spacing make it challenging to estimate the Fundamental Diagram (FD), which limits the utility of FD in determining the optimal platoon size at the macroscopic level. To bridge this gap, this paper theoretically analyzes the dynamics of platooning with the corresponding FD under the impacts of lane-changing maneuvers in multi-string environments. First, we develop a probabilistic model to characterize the distribution of platoon sizes as a function of lane-changing frequency, thereby quantifying the disruptive impact of lane changes on platoon formation. Second, to account for the additional spacing introduced by lane-changing maneuvers, we propose the concept of effective spacing based on minimum-energy principles, from which we derive the relationship between traffic flow and cruising speed. Third, we investigate the impact of lane-changing frequency and the maximum allowable platoon size on the shape and evolution of the FD. Theoretical derivations and simulation results are used to support platoon management in multi-string environments by providing quantitative insights into how lane-changing behaviors influence traffic flow and overall roadway capacity.

4 - Evaluating the Impacts of Mental Stress on Car-Following Performance Under Cav Cyberattacks in Mixed-Flow Traffic

Yangjiao Chen, Georgia Institute of Technology, Atlanta, GA, United States, Shubham Agrawal, Irina Benedyk, Srinivas Peeta

As connected and automated vehicles (CAVs) become increasingly prevalent, their susceptibility to cyberattacks introduces new safety challenges. While analytical methods for detecting such threats are advancing, there is little literature on the psychological effects of these disruptions on human drivers. Given the rarity and unpredictability of real-world CAV cyberattack scenarios, this study employs interactive, simulator-based experiments to investigate human responses when following compromised CAVs. A high-fidelity, human-in-the-loop driving simulator—integrating real-world SAE Level 4 vehicle trajectories—is employed to assess the effects of ongoing cyberattacks on car-following behavior. The simulator is equipped with in-vehicle driver monitoring systems that capture multimodal physiological data, including eye-tracking, electrocardiography (ECG), and electroencephalography (EEG), to assess mental stress. Eye-tracking features (e.g., fixation patterns, blink rates), heart rate variability (HRV) derived from ECG, and changes in brain activity from EEG, are analyzed. These physiological indicators are validated against behavioral measures (e.g., reaction time) and subjective assessments (e.g., post-drive questionnaires). Car-following performance is quantified using metrics including mean headway distance, headway variance, and root-mean-square (RMS) velocity error. The results indicate significant changes in driving performance when drivers perceive CAVs to be compromised by cyberattacks. These findings provide actionable insights for the development of adaptive vehicle interfaces and mitigation strategies that enhance driver safety and performance under cyber-physical disruptions.

5 - Human-like Lane-Change Control Strategy for Connected and Autonomous Vehicles to Improve Interactions with Human-Driven Vehicles

Yongyang Liu, Georgia Institute of Technology, Atlanta, GA, United States, Srinivas Peeta

Connected and autonomous vehicles (CAVs) promise enhancements to traffic safety, efficiency, and smoothness. However, their coexistence with human-driven vehicles (HDVs) in mixed-traffic environments presents significant challenges such as weakened CAV control performance and induced mental discomfort for HDV drivers, particularly during lane changes. To address these challenges, this study proposes a human-like lane-change control strategy (HLCS) for CAVs to integrate human intelligence and cognition with advanced CAV capabilities (in situational awareness and reaction times) to produce human-like lane-change behavior, improving CAV-HDV interactions and promoting smooth integration of CAVs into existing traffic flow. Operating within a theory of mind framework, the HLCS introduces human-like reasoning into CAV operations. It first develops a human-like lane-change model (HL-LCM) that combines the enhanced control of robot driving and the human expertise of natural human driving. The model produces human-like lane-change maneuvers to improve CAV performance and mitigate HDV drivers' mental discomfort. Further, the HLCS proposes a HDV behavior predictor that learns human driving patterns to model human decision-making. By anticipating HDV responses to CAV lane changes, the HL-LCM performs informed and adaptive maneuvers that mitigate disruptions and maintain traffic smoothness. Numerical experiments illustrate that HLCS can leverage rich human driving experience and advanced CAV capabilities to improve CAV-HDV interactions in lane changes, resulting in fewer traffic oscillations and enhanced safety, CAV smoothness, and mental comfort for HDV drivers. The study provides insights for developing human-like CAV systems to manage complex interactions with HDVs, to promote human-friendly mixed-traffic environments.

6 - Revisiting Followerstopper: Phase-Space Based Traffic Decongestion Controller

Rahul Bhadani, The University of Alabama in Huntsville, Huntsville, AL, United States

We provide a detailed description of the control law in the Followerstopper, a velocity controller based on the phase-space, that demonstrated its ability to mitigate emergent traffic jams due to stop-and-go traffic during rush hour in the mixed-autonomy setting. We deployed the Followerstopper on an autonomous vehicle to follow human drivers. The controller attenuates the emanant traffic waves by regulating its velocity according to the relative distance and velocity of the leader car. While controlling the velocity, the controller also prevents the collision of the ego vehicle with the lead vehicle within the range specified by the controller's design parameter. The controller design is based on the relative distance-relative velocity phase-portrait that allows the transition of the regulated velocity from (i) no modification of the desired maximum velocity, (ii) decelerating to match the leader's velocity; (iii) braking to avoid an imminent collision. Further, we explore the phase-space properties of the Followerstopper and provide a detailed description of a nonlinear control law that regulates the reference input to the Followerstopper within the physics-informed boundaries to prevent any unsafe behavior or imminent collision.

TC13

Building A Level 3 A315

Sustainable and Resilient Transportation Systems

Contributed Session

Contributed

Chair: Prajwal Chauhan, A4-303, 129188

1 - Fresh Produce Spatial Price Equilibrium on General Networks: Capturing Commodity Quality Deterioration Through Endogenous Transportation TIME Delay Functions with Capacities

Samirasadat Samadi, University of Massachusetts Amherst, Amherst, MA, United States, Anna Nagurney, Deniz Besik

Fresh produce is vital to global food security but is highly vulnerable to transportation delays and capacity disruptions, as in critical links such as the Panama Canal. This paper develops a multicommodity fresh produce trade equilibrium model under quality deterioration on a general network, allowing for multiple transportation links on a path between supply and demand markets. We extend the Bureau of Public Roads (BPR) congestion function to endogenize time delay as a function of all commodity shipments on a link. By embedding link upper bounds in time delay functions, we capture how upper bounds impact transportation times and fresh produce quality. The spatial price equilibrium conditions are formulated as a variational inequality problem. We also propose commodity quality trade network performance measures — supply-based, demand-based, and network-based — that can be applied for an individual commodity or across all commodities. We apply the model to bananas, a globally popular and nutritious fruit, and its trade from Ecuador and Costa Rica to the European Union, the United States, and Russia, focusing on drought-driven congestion at the Panama Canal. Numerical scenarios show that reductions in transportation link upper bounds significantly lower banana shipments, degrade their quality, and shift prices at both supply and demand markets. The results align closely with real-world data on banana prices, transportation costs, and export volumes, reinforcing the model's

practical relevance. Our model provides decision-makers with essential insights on the impacts of congestion and increasing transportation times on fresh produce trade and its quality, along with the performance of the trade network.

2 - Generative AI-Driven Stochastic Maritime Route Optimization Under Environmental Uncertainty for Fuel Efficiency

Hyunha Lee, Pusan National University, Busan, Korea, Republic of, Dohee Kim, Yongjae Lee, Hyerim Bae

As maritime trade volume continues to grow, carbon emissions from the shipping industry are increasing accordingly. To address this challenge, the International Maritime Organization (IMO) has implemented stricter environmental regulations, including the Carbon Intensity Indicator (CII) and the Ship Energy Efficiency Management Plan (SEEMP). These developments necessitate transitioning from voyage planning focused on punctuality to strategies considering fuel efficiency and operational uncertainty. Prior research has predominantly employed deterministic approaches, which are limited in capturing uncertainties like route deviations caused by regional heterogeneity or environmental changes. While recent studies have introduced a stochastic generative artificial intelligence model, their disregard for fuel efficiency limits their usefulness in sustainable operations planning. To fill this gap, this study presents a data-driven framework that builds a dynamic maritime network using regional ocean characteristics such as congestion, environmental conditions, and interregional connectivity. A probabilistic route generation process is then employed to simulate feasible routes within the network. Each candidate route is assessed based on predicted fuel consumption, and the optimal route is chosen to minimize expected fuel consumption while adhering to punctuality constraints.

The proposed method is validated by using real-time sensor data from the Automatic Identification System (AIS) and Metocean datasets. The results demonstrate its effectiveness in generating operationally feasible routes optimized for fuel efficiency under uncertainty. Integrating stochastic modeling and predictive analytics into maritime network-based route optimization facilitates sustainable logistics practices aligned with environmental regulations. It also contributes to decision support systems that advance decarbonization in the shipping industry.

3 - Underground Freight Transportation for Middle Mile Delivery

Sarah Powell, University of St. Thomas, St. Paul, MN, United States, Ann Campbell

The use of underground tunnels for freight transportation is gaining attention in both industry and academia as a way to efficiently move packages between or within urban areas. We propose a new problem, the middle mile tunnel dispatching problem (MMTDP), to capture the use of tunnels between cities, inspired by the Cargo Sous Terrain project in Switzerland. We model this problem as an integer program with the objective of maximizing the demand volume that arrives at its destination by its primary deadline. In the MMTDP, vehicles carry pallets of goods between distribution centers in different cities. We evaluate the complexity of different variants of the MMTDP and provide a set of computational results to demonstrate the tunnel system's performance and the impact of different design parameters. We find that the size of tunnel vehicles significantly impacts the amount of demand that can be delivered by its deadline and a larger proportion of tunnel capacity can be utilized when there are more orders with short distances between origin and destination pairs. We offer managerial insights on dispatching strategies for linear middle-mile tunnel networks. For example, we show that it can be helpful to consolidate orders with the same origin, especially for small tunnel vehicle sizes, and the same origin and destination with larger tunnel vehicle sizes.

4 - Weather Impact on Fast Charging Infrastructure Planning for Large-Scale Electric Vehicle Fleets (Archive Only)

Gustavo Campos, National Renewable Energy Laboratory, Golden, CO, United States, Roberto Vercellino, Yanbo Ge, Monte Lunacek, Devon Sigler, Juliette Ugurumera, Margaret Mann

Fast charging technology is an essential component for promoting the rapid adoption of electric vehicles. Although shown to have a significant impact, weather conditions - particularly low temperatures - are often neglected in the literature due to the lack of experimental data. This work aims to fill this gap by investigating the impact of weather conditions or low-temperatures on fast charging infrastructure planning, particularly in the context of high-utilization fleets, such as airport rental car facilities, which may handle up to thousands of vehicles per day with quick turnaround times. A recently developed discrete event simulator is used to represent the operations of a large-scale electrified rental car facility, while a recently published experimental dataset is used to incorporate weather effects on fast charging curves. Different ways to perform the latter are discussed. Results demonstrate the impact of weather on infrastructure planning metrics, such as number of chargers, customer wait times, and station peak loads.

5 - Modeling Early-Stage Advanced Air Mobility for Commuting with Integrated Air-Ground Services

Bo Zou, University of Illinois Chicago, Chicago, IL, United States, Heeseung Shon, Danesh Hosseinpahani

This study proposes a mathematical framework for optimizing an Advanced Air Mobility (AAM)-embedded multimodal transportation system at an early-stage, focusing on coordinated operations between electric vertical take-off and landing aircraft (eVTOLs) and shuttle services for commuting between a suburb and the central business district in a metropolitan area. The framework addresses joint scheduling of air and ground vehicles to ensure first- and last-mile connectivity and meet commuter demand, as well as eVTOL charging and relocation to balance AAM supply and demand. An integer linear program (ILP) is formulated on a time-expanded network encompassing both air and ground operations. Given the numerous variables and constraints involved and the combinatorial nature of the ILP, commercial solvers can yield optimal solutions only for small instances. Recognizing the fact that the ILP involves only a few coupling constraints, a dual decomposition solution approach based on the Alternating Direction Method of Multipliers (ADMM) algorithm is devised. The approach decomposes the ILP into eVTOL operation, shuttle operation, and passenger routing subproblems. A subproblem is decomposed further into even smaller subproblems, each dealing with the routing of just an individual eVTOL, shuttle, and passenger group. We implement the ILP and the solution approach in the Chicago metropolitan area. The numerical results provide several insights into the practical implementation of integrated AAM systems and strategies to enhance the system operational efficiency and passenger satisfaction. The proposed solution approach is also found to significantly reduce the computation time while achieving near optimal solutions.

6 - The Geography of Transportation Cybersecurity: Visitor Flows, Industry Clusters, and Spatial Dynamics

Kailai Wang, University of Houston, Houston, TX, United States, Yuhao Wang

The rapid evolution of the transportation cybersecurity ecosystem—encompassing cybersecurity, automotive, and transportation and logistics sectors—has led to the formation of distinct spatial clusters and visitor flow patterns across the US. This study examines the spatiotemporal dynamics of visitor flows, analyzing how socioeconomic factors shape industry clustering and workforce distribution within these evolving sectors. To model and predict visitor flow patterns, we develop a BiTransGCN framework, integrating an attention-based Transformer architecture with a Graph Convolutional Network (GCN) backbone. Using SafeGraph's 2022 weekly visitor flow data along with additional socioeconomic indicators, we conduct a multi-scale spatial clustering analysis to identify distinct industry concentration trends, workforce mobility patterns, and geographic interdependencies across three key transportation cybersecurity-related industries. Our findings reveal that geolocation and education levels are the most significant factors influencing industry cluster formation. Furthermore, our predictive model anticipates a 14.16% average increase in visitor flow across the US in the next term, highlighting emerging industry growth, regional workforce shifts, and cross-sector innovation. By integrating AI-driven forecasting methods with spatial analytics, this study enhances our ability to track, understand, and anticipate shifts in industry clustering and mobility trends, supporting strategic planning for a secure and resilient transportation network. This research provides a data-driven foundation for economic planning, workforce development, and targeted investments in the transportation cybersecurity ecosystem.

7 - Neural Operators Struggle to Learn Complex Pdes in Pedestrian Mobility: Hughes Model Case Study

Prajwal Chauhan, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates, Salah Eddine Choutri, Mohamed Ghattassi, Nader Masmoudi, Saif Eddin Jabari

This work investigates the limitations of neural operators in learning solutions for the Hughes model, a first-order hyperbolic conservation law for crowd dynamics. The model couples a Fokker-Planck equation representing pedestrian density with a Hamilton-Jacobi-type (eikonal) equation. The Hughes model belongs to the class of nonlinear hyperbolic systems that often exhibit complex solution structures, including shocks and discontinuities. In this study, we evaluate the performance of three state-of-the-art neural operators (Fourier Neural Operator, Wavelet Neural Operator, and Multi-wavelet Neural Operator) in various challenging scenarios. Specifically, we consider discontinuous and Gaussian initial conditions with diverse boundary conditions, while also examining the impact of different numerical schemes.

Our results show that these neural operators perform well in easy scenarios with fewer discontinuities in the initial condition, but struggle in complex scenarios with

multiple initial discontinuities and dynamic boundary conditions, even when trained specifically on such complex samples. The predicted solutions often appear smoother, resulting in a reduction in total variation and a loss of important physical features. This smoothing behavior is similar to what happens in models that introduce artificial diffusion, causing them to miss essential features, such as shock waves, in hyperbolic systems. These results suggest that current neural operator architectures may introduce unintended regularization effects that limit their ability to capture transport dynamics governed by discontinuities. They also raise concerns about generalizing these methods to traffic applications where shock preservation is essential.

TC14

Building A Level 3 A316

EV Fleet Optimization and Operations

Contributed Session

Contributed

Chair: Hyeon Suk Na, University of Missouri, Columbia, MO, United States

1 - Large-Scale Evacuation with Electric Vehicles: Emerging Challenges in South Texas Disaster Management

Joon-Yeoul Oh, Texas A&M University-Kingsville, Kingsville, TX, United States, Olusola Ogungbeje, Delia Valles-Rosales, Nuri Yilmazer

The expanding adoption of electric vehicles (EVs) in the U.S. is transforming transportation by reducing reliance on fossil fuels and curbing emissions. However, concerns remain regarding vehicle range limitations, the vulnerability of the power grid during disasters, and gaps in charging infrastructure, all of which may compromise transportation reliability during large-scale evacuations for hurricanes, wildfires, or other emergencies. There is little evidence regarding the security, sustainability, and accessibility of EV charging stations during such events. The proposed study will estimate the number of EVs in the South Texas area by 2035 and investigate the impacts of emergency evacuations involving EVs and their charging stations during natural disasters. This research aims to assess the sustainability and accessibility of EVs during such events and to propose strategies for increasing accessibility in the South Texas region. The study will assist stakeholders in designing electric vehicle systems, including disaster-resilient infrastructure for EVs and charging stations, to mitigate risks and ensure safer, more inclusive evacuation strategies.

2 - Modeling and Solving the Electric Truck and Trailer Routing Problem Using Simulated Annealing with Extensions to Other EV Routing Variants

Ahmad Bassaleh, Wichita State University, Wichita, KS, United States, Mehmet Yildirim

The integration of electric vehicles (EVs) into logistics systems has accelerated in recent years, motivating the development of advanced routing models that account for energy limitations, charging times, and customer time windows. This study extends the classical Electric Vehicle Routing Problem with Time Windows (EVRPTW) to a variant based on the Truck and Trailer Routing Problem (TTRP). In this setting, trucks may carry detachable trailers that must be left behind and picked up later when visiting specific customers. We propose a novel mixed-integer linear programming (MILP) formulation that captures the operational constraints of electric trucks, including energy consumption, charging station usage, time window compliance, and trailer-handling logic. We develop a customized Simulated Annealing (SA) algorithm tailored to the problem structure. For benchmarking, we modified existing EVRPTW instances to incorporate trailer detachment constraints. The SA algorithm was applied to these extended instances, and its performance was evaluated through comparative analysis. To assess the generalizability of our approach, the same SA algorithm was also applied to standard EVRPTW instances. Results demonstrate that our method is competitive with existing solutions while offering greater modelling flexibility. This adaptability makes both the model and algorithm well suited to a broad range of electric vehicle routing scenarios, supporting the advancement of sustainable transportation systems.

3 - Two-Stage Stochastic EV Fleet Planning for E-Commerce Delivery via Predictive Learning

Manzi Li, University of Michigan, Ann Arbor, MI, United States

The escalating demand in e-commerce, coupled with the global push for sustainability, has made the operational planning of electric vehicle (EV) fleets a problem of critical importance. The inherent uncertainties in daily customer demand and potential charging station congestion make this a complex challenge. To capture this complexity, we formulate the EV fleet deployment and routing problem as a two-stage stochastic optimization model. The first-stage decisions determine the strategic fleet size and vehicle-to-depot assignments, while the second-stage recourse actions define the operational routing and charging schedules once demand is realized. However, the computational burden of solving such models prohibits their use in real-time decision-making environments. To overcome this, a novel machine learning framework is introduced in this paper. We first generate a comprehensive training dataset by solving numerous problem instances offline. A predictive model is then trained to learn the direct mapping from problem parameters to the optimal here-and-now decisions. In the online phase, these predictions are used to formulate a simplified, single-level optimization problem for the wait-and-see actions, which can be solved almost instantaneously. This approach transforms a computationally intractable problem into a practical, data-driven tool for dynamic and sustainable e-commerce logistics.

4 - Optimizing Fleet Composition and Charging Strategies for Urban Bus Systems

Sadjad Bazarnovi, University of Illinois Chicago, Chicago, IL, United States, Taner Cokyasar, Omer Verbas, Abolfazl Mohammadian

Electrification of transit fleets offers a potential pathway to reducing cost of mobility. However, the limited battery range and long charging durations of battery electric buses (BEBs) introduce significant operational complexities, necessitating innovative scheduling and charging strategies. This study proposes an integrated Mixed-Integer Linear Programming (MILP) model to optimize vehicle scheduling and charging strategies for mixed fleets of BEBs and Diesel Buses (DBs). Unlike existing models, which often assume a fixed BEB fleet size or restrict charging to a single charger type, our approach simultaneously determines the optimal fleet composition, scheduling, and flexible and partial charging strategy incorporating both slow and fast chargers at garages and terminal stations. The model minimizes combined fleet purchase and operational costs. A deterministic queueing strategy is introduced, departing from traditional first-come, first-served methods by dynamically allocating waiting and charging times based on operational priorities and resource availability, improving the overall scheduling efficiency. To overcome computational complexities arising from numerous variables, a genetic algorithm-based metaheuristic (GA) is developed, facilitating scalable solutions for large-scale transit networks. Numerical experiments using real-world transit data from the Chicago Transit Authority and the Pace suburban bus systems demonstrate the model's effectiveness, highlighting that BEBs, despite higher initial acquisition costs, effectively serve a greater proportion of trips due to lower energy costs. Results highlight the feasibility and economic viability of mixed-fleet electrification, illustrating the potential for substantial operational efficiencies through optimized scheduling and flexible charging strategies.

5 - Integrated Bus Fleet Electrification Planning Through Accelerated Logic-Based Benders Decomposition and Restriction Heuristics

Robin Legault, Massachusetts Institute of Technology, Cambridge, MA, United States, Filipe Cabral, Xu Sun

To meet sustainability goals and regulatory requirements, transit agencies worldwide are planning partial and complete transitions to electric bus fleets. This work presents a comprehensive and computationally efficient multi-period optimization framework integrating the key planning decisions necessary to support such electrification initiatives. Our model, formulated as a two-stage integer program with integer subproblems, jointly optimizes yearly fleet and charging infrastructure investments as well as hourly vehicle scheduling and charging operations. To solve instances of practical relevance to proven optimality, we develop a logic-based Benders decomposition method enhanced by several techniques, including preprocessing, partial decomposition, and a range of classical and monotone Benders cuts derived from relaxations of the operational subproblems. These accelerations yield speedups of up to three orders of magnitude, and lead to practical and theoretical insights into Benders cut selection. We also propose a heuristic tailored for long-term, citywide electrification planning. This approach, which imposes and progressively relaxes additional scheduling constraints, consistently delivers high-quality solutions with optimality gaps below 1% for instances an order of magnitude larger than those considered in prior studies. We illustrate our model using data from the Chicago public bus system, providing managerial insights into optimal investment and operational policies.

6 - Commercial Megawatt Charging Reservation Mechanism

Ridvan Aksu, The University of Alabama, Tuscaloosa, AL, United States, Goutham Takasi, Mesut Yavuz

Long-haul electric freight trucks require multiple charging stops in public stations, risking hours of waiting time in an industry governed by strict schedules. Megawatt charging stations (MCS), while capable of rapid energy delivery, place significant strain to the power grid, thus has to operate at a limited capacity. To address this challenge, we propose a pricing mechanism design for MCS that allows electric trucks with tighter schedules to reserve charging slots in advance. We demonstrate that an optimal pricing mechanism exists and allows EV trucks to alleviate long waiting times, while keeping charging stations balanced and profitable.

7 - Post-Disaster Damage Assessment from High-Resolution Satellite Imagery Using D-Linknet Hierarchical Transformer with Radon-Transformed Fixed Positional Encoding

Hyeong Suk Na, University of Missouri, Columbia, MO, United States, Shivam Shrivastava

Accurate and timely damage assessment is critical for effective emergency response and post-disaster recovery. A key task in this process is the generation of high-resolution damage segmentation maps that evaluate various levels of structural damage. Typically, this task includes three steps: feature extraction, change detection, and damage segmentation. Existing studies utilize these three steps independently on pre- and post-disaster satellite imagery to assess the damage level of buildings regionally. However, decoupling these steps can limit the evaluation speed and accuracy of damage assessment. To consider this gap, we propose a deep learning architecture, D-Linknet Hierarchical Transformer with Radon-Transformed Positional Encoding (DLHT-RTPE), that integrates all three steps in a single framework. We first preprocess satellite imagery using a Gaussian filter to improve image sharpness to facilitate effective feature extraction. The enhanced images are then processed through a dilation module to preserve high-level semantic information in downsampled, low-resolution feature maps. The change detection is performed at each level of feature maps using the hierarchical transformer-based difference blocks optimized with RTPE, enabling precise segmentation of damage at the highest resolution of feature maps. The RTPE method captures the sinusoidal patterns inherent in imagery and integrates them into the transformer decoder to recover the original token positions. This approach significantly reduces the number of learnable parameters and achieves a 10% reduction in prediction time. The proposed DLHT-RTPE model, trained on the xBD dataset, achieved a comparable accuracy of 81.5%, demonstrating efficient performance and computational efficiency compared to state-of-the-art models such as DAHiTrA and BiTransUNet.

TC15

Building A Level 4 A401

Machine Learning and Optimization in Aviation

Contributed Session

Contributed

Chair: Abhilash Vimal, The University of Texas at Arlington, 701 West Street, Arlington, TX, 76019, United States

1 - Baggage Handling via Bandit-Augmented Reinforcement Learning

Keyu TIAN, City University of Hong Kong, Hong Kong SAR, Hong Kong, Li Zeng

Efficient baggage delivery within baggage handling system (BHS) at airport is critical for optimal airport operations. Traditional shortest-path based routing planning methods, which depend on fixed strategies, often struggle to adapt to the complex, dynamical conditions of BHS and require frequent re-planning to avoid potential baggage collisions. Reinforcement learning (RL), particularly deep double Q-networks (DDQN), has shown promise in addressing these routing challenges. However, current studies mainly focus on controlling diverters for optimal baggage routing, neglecting other influential factors such as the arrangement of entry (check-in counters) and exit (unloading areas) points for different airlines, which can significantly impact routing efficiency. Moreover, many RL-based approaches rely on simplified models that fail to capture the complexity of real-world system dynamics, leading to suboptimal performance. To address these issues, we propose a system-level framework for baggage handling that simultaneously optimizes diverter control and airline entry-exit arrangement to enhance performance. This framework employs a DDQN-based routing strategy with adaptive learning for diverter control and uses the upper confidence bound (UCB), a bandit method, for airline entry-exit selection. Furthermore, a kernel-based data generation method is incorporated to create more realistic and diverse baggage data in order to improve the robustness and generalization of the learned policy. Experiments show that the proposed method outperforms existing approaches in accelerating baggage delivery and preventing collisions.

2 - Bridging the Revenue Gap: a Hierarchical Bayesian Framework for Personalized Airline Upgrade Pricing

Peng Li, Embry-Riddle Aeronautical University, Daytona Beach, FL, United States, aqi dong, Farshid Azadian, Lin Zhao

Airlines face a critical challenge in traditional tier-based upgrade pricing: capturing customers' true willingness-to-pay while accounting for individual preferences. Current approaches achieve only 68% of potential revenue due to rigid pricing structures that fail to adapt to individual customer contexts. This research introduces a hierarchical Bayesian framework that addresses this problem by modeling passenger willingness-to-pay at both individual and population levels. The methodology creates an adaptive personalization mechanism that balances sparse individual-level data with robust population-level information, enabling precise price customization even for customers with limited historical data. Our approach incorporates explicit uncertainty quantification in pricing decisions, allowing risk-aware pricing strategies that maximize expected revenue. Furthermore, we model contextual effects such as trip purpose, booking timing, and competitive offerings that significantly influence purchase decisions. Implementation results demonstrate 91% revenue capture compared to theoretical optimum—a 34% improvement over traditional approaches. The framework provides airlines with an implementable solution for dynamic personalized pricing that balances revenue maximization with computational feasibility, while respecting customer expectations regarding fair pricing practices.

3 - The Price of Routing with Preferences for Uncrewed Aerial Vehicles

Geoffrey Ding, Massachusetts Institute of Technology, Cambridge, MA, United States, Hamsa Balakrishnan

Increasing operations of uncrewed aerial vehicles (UAVs), or drones, bring growing concerns about safety, privacy, and noise. These risks may be reduced by routing along paths with less human activity such as waterways and railways, which we call "preferred segments." By varying the number of preferred segments and the degree of their preference in a modified visibility graph, we show how preferences can affect routing and operational efficiency. We quantify these impacts as the "price of preference." Illustrative examples of these concepts are provided in a synthetic environment and a drone-as-first-responder (DFR) case study in Detroit, Michigan. This work provides one possible approach for balancing the efficiency of drone operations with other public interests.

We adapt the weighted region problem to address path planning with preferred segments. By varying the degree to which preferred segments are discounted, we find a corresponding change in the efficiency of the resulting paths. We assess how different discount factors influence how much operations follow preferred segments in obstacle-free settings. We then extend visibility graphs, which allow us to model obstacles or no-fly zones, to account for preferred segments. Obstacles may alter the efficient path, so we also consider how they may change the lowest-cost path and the price of using preferred segments. Illustrative examples of these concepts are provided, including a case study of medical deliveries in Detroit, Michigan. This work presents one possible framework for trading off efficient operations of novel aircraft with public objectives such as safety and privacy.

4 - Sustainable Aviation Fuel (SAF) Supply Chain Design (in Thailand)

Rakesh Nagi, Singapore University of Technology and Design (SUTD), Singapore, Singapore, Jaghpaal Singh Baram

Abstract

Growing global air travel demands an important shift towards Sustainable Aviation Fuel (SAF) to lessen its environmental impact. This research develops a robust supply chain modeling and optimization framework for the entire SAF lifecycle, from feedstock sourcing to final airport delivery. Our core objective is to identify strategic opportunities to streamline processes, substantially reduce overall costs, minimize environmental footprints, and determine optimal biorefinery locations throughout the chain. The study's specific focus is Thailand, a nation with abundant, year-round availability of agricultural residues. Utilizing these waste-to-value feedstocks provides an excellent opportunity of transforming waste and reducing aviation environmental footprint at the same time!

The framework considers two primary conversion pathways that appear to be promising options: Biomass-to-Liquid Fischer-Tropsch (BtL-FT) synthesis and Cellulosic Alcohol-to-Jet (ATJ). These pathways involve unique logistical challenges like bulky residue handling, processing complexities, and significant capital (CAPEX) and operational (OPEX) expenditures. Thailand's consistent feedstock supply is pivotal, as its year-round availability critically enhances the viability of continuous refinery operations, especially for the capital-intensive BtL-FT process. A central component of this research involves applying supply chain modeling and optimization to strategically determine optimal biorefinery locations across Thailand. This meticulous planning is designed to minimize transportation routes, associated costs, and logistical constraints across the entire supply chain. A comprehensive mathematical model will be introduced to formalize and solve this complex, multi-objective optimization problem. Ultimately, this study aims to significantly enhance SAF's accessibility and affordability through data-driven strategic planning, fostering a greener aviation industry for sustainable future generations.

5 - Balancing Exploration and Exploitation: How Resource Reconfiguration and Sustainability Orientation Influence Market Performance in Global Aviation

Abhilash Vimal, The University of Texas at Arlington, Arlington, TX, United States, Seock-Jin Hong, Edmund Prater

The present manuscript empirically investigates the role that a firm's resource reconfiguration ability and sustainability orientation play in influencing the relationship between their ambidexterity efforts and market performance. Ambidexterity is further diverged into exploration and exploitation orientation. Data for the study comes from a survey administered to airline employees and airline customers of 13 US and Canadian passenger airlines. The commercial airline industry was an appropriate base for the study because of its high overhead/capital costs, sensitivity to jet fuel prices, susceptibility to disruption which are further aggravated by macro-predictability, micro-uncertainty & capacity constraints. Structural equations modeling is used for this study. We assess the positive association between an airline's exploration & exploitation orientation on market performance to be weaker for lower levels of resource reconfiguration ability. Likewise, the positive association between an airline's exploration/exploitation orientation and market performance is stronger at higher levels of sustainability orientation. Most significantly, the moderating effects of resource reconfiguration and sustainability orientation is stronger for a strategy of ambidexterity which combines exploitation and exploration orientation. Despite the abundance of research on the relationship between efficiency and performance, there is a dearth of research that deciphers the factors that affect ambidexterity outcomes for passenger airlines as extant literature has predominantly examined the effect of technical efficiencies, cost efficiencies and environmental efficiencies. This study sheds more light on the role of resource realignment and sustainability orientation on ambidexterity outcomes. Additionally we embed ambidexterity within a richer contextual framework by incorporating strategic alignment with environmental values and adaptive capabilities.

TC17

Building A Level 4 A403

AI for Energy Systems

Invited Session

Practice Track

Chair: Russell Bent, Los Alamos National Laboratory, Los Alamos, United States

1 - Harnessing Artificial Intelligence for Enhanced Efficiency in Energy Systems: Challenges and Opportunities

Femi Omिताomu, Oak Ridge National Laboratory, Oak Ridge, TN, United States, Teja Kuruganti

The integration of artificial intelligence (AI) into energy systems represents a transformative evolution in the planning, deployment, and operational phases of energy management. This talk will explore the practical applications of AI technologies within the energy sector, with a specific focus on generation planning, site selection, demand forecasting, and operational optimization. By leveraging advanced algorithms and machine learning techniques, AI enhances decision-making processes, providing accurate predictions and insights that improve resource allocation and efficiency in generation planning. Furthermore, AI plays a crucial role in site evaluation by analyzing geographic and economic factors to identify the most advantageous locations for energy generation assets. The focus is on the importance of AI in ensuring optimal energy supply, as well as its impact on enhancing the operational performance of existing assets through predictive maintenance and resilience evaluations. Despite the promising advantages that AI offers, its implementation in the energy sector faces several challenges. Issues such as data integration, privacy concerns, algorithm trustworthiness, and workforce development must be addressed to fully harness the potential of AI technologies. By examining these aspects, this talk aims to provide a comprehensive understanding of how AI can revolutionize energy systems while also highlighting the necessary steps needed to overcome barriers to successful deployment.

2 - Bottleneck Challenges in Deploying AI to Power Systems Today

Colin Ponce, Lawrence Livermore National Laboratory, Livermore, CA, United States

The technology industry that produced AI's recent groundbreaking innovations often operates by "moving fast and breaking things." While this approach to technology development can lead to rapid progress, its acceptance of risk does not translate well to critical infrastructure sectors such as the power industry, where outages can be highly disruptive and even life-threatening. Due to the high cost of outages, the power industry is often resistant to deploying a new technology unless (A) it integrates seamlessly with the existing power grid and all its legacy systems, and (B) those deploying it have confidence that it will work as intended. To enable the power industry to leverage the massive potential of AI, therefore, we cannot focus only on the task of developing useful AI technology. The key challenge in this space is rather in creating a deployment pathway that meets the dual requirements of integration with existing systems and ability to ensure confidence in operation. In this talk, I will present current progress and pathways towards meeting these industry needs. First, I will discuss how in-use AI applications are both integrable and trustworthy, and I will show how some in-progress projects are building these requirements into every aspect of their development. Second, I will discuss some emerging approaches, such as AI testbeds and explainable AI, that can assist developers in meeting the requirements of integrating AI into power systems.

3 - The Role of AI in Control and Optimization of Energy Systems

Andrey Bernstein, NREL, Golden, CO, United States

Modern power and energy systems are rapidly evolving into complex, networked systems with significant challenges in design, planning, and operation across different levels. At the transmission level, a major challenge lies in the real-time monitoring and control of numerous dynamic "what if" scenarios. At the distribution level, distributed energy resources (DERs) are integrated at the system edge, often not visible to the system's operators. The future power system is anticipated to be a networked system of energy communities managed automatically, and in emergencies, semi-automatically, aggregating DERs at the customer level and functioning as virtual power plants (VPPs). This transformation renders traditional, model-based power system planning and operation paradigms inadequate, as they largely ignore variability and uncertainty at the distribution level. Precise physical models of complex energy systems are often unavailable, making model-based optimization and control approaches infeasible. Furthermore, as more controllable devices are added at the system's edge, human behavior increasingly influences the system, further complicating traditional planning and operation processes. These challenges highlight the need for the careful integration of AI/ML-based approaches in power and energy system planning and operation, with a strong emphasis on physics-informed AI/ML to better leverage physical system information. This talk provides an overview of major research challenges in power and energy systems that cannot be addressed using classical methods; the landscape of AI/ML methods suitable for energy system applications; examples of successful AI/ML applications; and an overview of open problems and the potential of AI/ML-based methods to solve them.

TC19

Building A Level 4 A405

Top-Notch Applications of Queueing Theory

Invited Session

Applied Probability Society

Chair: Daniela Hurtado Lange, Northwestern University, Evanston, IL, United States

1 - Efficiency-Cost Trade-off in Queues with Dynamic Arrivals

Sushil Mahavir Varma, INRIA Paris, Paris, France

With the information technology revolution, online marketplaces like ride-hailing platforms and payment channel networks have become integral to our society. In such systems, system operators need to design optimal policies to improve customer experience by minimizing delays (efficiency) while optimizing platform rewards (cost), such as profit in ride-hailing and the number of successful transactions on payment channel networks. However, these dual objectives often conflict, naturally leading to a trade-off. In this talk, we characterize this fundamental trade-off observed in service systems.

We consider an M/M/1 queue with state-dependent arrivals, where the system operator can adjust the arrival rate $\lambda(q)$ as a function of the queue length. The service rate is fixed at 1. The efficiency is measured by $E[q]$, and the system operator's reward is defined by a general function F , given by $E[F(\lambda(q))]$, which represents the expected reward in the steady state. Let F^* denote the maximum attainable reward and define the cost as $F^* - E[F(\lambda(q))]$. The objective is to analyze the cost-efficiency trade-off: for a given $\epsilon > 0$, if cost $\leq \epsilon$, then, we show that $E[q] \geq d(\epsilon)$ under any state-dependent arrival setting, where d depends only on the reward function F . Contrary to the classical $1/\epsilon$ scaling for the static arrival setting, we show that $d(\epsilon) = \Omega(1/\epsilon^{0.5})$ when F is strictly concave and $d(\epsilon) = \Omega(\log(1/\epsilon))$ when F is sufficiently non-concave. We show that these bounds are attainable by constructing a state-dependent arrival sequence $\lambda(q)$ achieving the minimum possible $E[q]$, characterizing the fundamental trade-off between cost and efficiency.

2 - Evaluating Relational Continuity in Primary Care: a Mean-Field Queueing Model

Yueyang Zhong, London Business School, London, United Kingdom

Relational continuity (RC) in primary care improves health outcomes but creates operational inefficiencies. We analyze this trade-off using a queueing model where patients are routed to either their regular physician (preserving RC) or the earliest available physician (reducing wait times). Through mean-field analysis, we derive closed-form solutions for the system-wide optimal RC level, proving that full continuity is suboptimal; instead, an intermediate balance minimizes total costs (waiting + increased visit frequency). Game-theoretic analysis of decentralized patient decisions reveals inefficient Nash equilibria (all-or-nothing continuity), highlighting the need for centralized control. Extending the model to heterogeneous patient classes and calibrating it with NHS data, we demonstrate that: (1) moderate RC—where patients occasionally forgo their regular physician—optimizes both clinical and operational outcomes; (2) the equilibrium-optimum gap is largest when visit/waiting costs are balanced or high-RC-need patients dominate; and (3) class-dependent policies tailored to patient needs outperform across-the-board continuity management. These results provide actionable insights for primary care policy design, showcasing how queueing theory and mean-field approximation can resolve healthcare operations dilemmas.

3 - Delay Announcement and Utility in Healthcare Queueing Models

Cameron Curtis, Northwestern University, Evanston, IL, United States

We analyze patient utility in a healthcare queueing model with delay announcements. We consider whether and when a patient should be given an estimate of their waiting time and, if they are given one, whether to provide an unbiased estimate or to inflate their expected wait. Patients are more sensitive to loss than gain (as is predicted in prospect theory) and are permitted to abandon. We present initial simulation results to estimate the optimal policy as well as preliminary analytical results.

4 - Dynamic Control of Stochastic Matching Systems in Heavy Traffic

Yaosheng Xu, University of Chicago, Booth School of Business, Chicago, IL, United States, Baris Ata

Bipartite matching systems arise in many service environments in which agents or tasks from two distinct sets must be paired dynamically under compatibility constraints. We study such systems in a stochastic and high-dimensional setting, where the diversity of match types introduces notable complexity. Our goal is to develop a dynamic control policy that maximizes the long-run discounted total reward by balancing matching gains, holding costs, and abandonment penalties. To address both analytical and computational challenges, we approximate the system using a Brownian control problem (BCP) formulated under a balanced network, which offers a convenient baseline and can be readily extended to unbalanced settings. We characterize the structure of the BCP solution and develop a deep learning-based approach to estimate the value function gradient. Building on this, we propose a practical matching policy that performs effectively in high-dimensional systems and demonstrates consistent improvements over existing policies in extensive numerical experiments.

5 - Design and Control of Large Service Systems with Delay-Dependent Service Times

Achal Bassamboo, Northwestern University, Evanston, IL, United States, Deniz Simsek, Ohad Perry

Service systems often exhibit complex dynamics when customer service times depend on the delay experienced in queue, a phenomenon observed across various settings such as healthcare and hospitality. This delay-dependent service time introduces non-Markovian dynamics that render exact analysis of large service systems intractable. Deterministic fluid models approximating the system dynamics uncover some operational challenges arising from delay-dependence. These systems can exhibit multi-stability, where the queue fluctuates between multiple equilibria. In large systems, these transitions become infrequent, potentially leading to prolonged periods of severe congestion referred as congestion collapse. Conversely, in smaller systems, the system frequently shifts between equilibria, causing high variability. To mitigate these issues, we propose two tailored strategies: For large systems, we introduce the “slingshot policy”, a dynamic control that alternates between first-come-first-served (FCFS) and last-come-first-served (LCFS) scheduling to stabilize the system at the more desirable equilibrium. For smaller systems, we propose a “depooling mechanism”, where the service pool is strategically divided into smaller, uni-stable subsystems, effectively eliminating multi-stability at the design stage. Our framework provides actionable solutions for system design and control. The proposed approaches are supported by both theoretical analysis and simulation studies, offering practical insights to improve the stability, reliability, and efficiency of service systems affected by delay-dependent dynamics.

TC20

Building A Level 4 A406

Control of Queues

Invited Session

Applied Probability Society

Chair: Hayriye Ayhan, ISyE Georgia Tech, Atlanta, GA, United States

1 - Dynamic Buffer Allocation in a Two-Stage Tandem Queue

David Kaufman, University of Michigan - Dearborn, Dearborn, MI, United States, Jairo Alejandro Martinez Correa, Gabriel Zayas-Cabán

Emergency department (ED) boarding—when admitted patients remain in the ED due to insufficient inpatient capacity—continues to drive ED overcrowding and degrade care delivery, underscoring the need for coordinated, system-level interventions. In this talk, we address the problem of dynamically adjusting buffer capacity after phase-one service in a two-stage tandem queueing model. The goal is to temporarily alleviate congestion in the first stage while minimizing total holding costs from jobs waiting or receiving service across both stages. We model this scenario using a finite-horizon continuous-time Markov decision process, whose optimality equations yield differential equations—distinct from the algebraic/recursive equations of infinite-horizon problems. After establishing the existence of a solution to these equations, we implement a time-discretization approach to approximate the optimal value function and corresponding policy. Numerical examples illustrate the structure of the discretized policy.

2 - Dynamic Server Assignment in Queueing Systems with Synergistic Servers and Abandonments

Bihan Chatterjee, Georgia Institute of Technology, Atlanta, GA, United States, Hayriye Ayhan, Sigrun Andradottir

We study tandem queueing systems in which servers work more efficiently in teams than on their own and customers are impatient in that they may leave the system while waiting for service. Our goal is to determine the server assignment policy that maximizes the long-run average throughput. We show that when each server is equally

skilled at all tasks, the optimal policy has all the servers working together at all times. We also provide a complete characterization of the optimal policy for Markovian systems with two stations and two servers when each server's efficiency may be task dependent. We show that the throughput is maximized under the policy which assigns one server to each station (based on their relative skill at that station) unless station 2 has no work (in which case both servers work at station 1) or the number of customers in the buffer reaches a threshold whose value we characterize (in which case both servers work at station 2). We study how the optimal policy varies with the level of server synergy (including no synergy) and also compare the optimal policy for systems with different customer abandonment rates (including no abandonments). Finally, we investigate the case where the synergy among collaborating servers can be task-dependent and provide numerical results.

3 - Control of Queues with Outsourcing

Mark Lewis, Cornell University, Ithaca, NY, United States

We consider the control of a single server with the opportunity to outsource some of the workload. Customers arrive according to a Poisson process and service times have a general distribution. There is a cost for each outsourced customer and holding costs for those in the system. We characterize the structure of the policy that minimizes the discounted and long-run average cost.

4 - Optimal Service Rate Control in Queues with Abandonments

Runhua Wu, Georgia Institute of Technology, ATLANTA, GA, United States

We consider a Markovian, infinite capacity single server queueing system where customers waiting in queue may get impatient and leave without being served. There is a cost associated with abandonments and a holding cost associated with customers in the system. We consider two variants of the model: customers pay a reward either at the time of arrival or at the time of service completion. The system manager can set the service rate of the server to a value in a compact set or idle the server and there is a service cost associated with each service rate. Under these assumptions, our objective is to characterize the service rate control policy that maximizes the infinite-horizon discounted reward and the long-run average reward for both variants of the model. We formulate the problem as a continuous time Markov decision process problem and show that for both variants and optimality criteria, the optimal service rate is non-decreasing in the number of customers in the system. We conclude with a study of finite capacity systems and show that the optimal service rate is not necessarily monotone in the number of customers in the system.

5 - Dynamic Pricing in Queueing Systems with Customer Abandonments and Reimbursement

Jieqi Di, Georgia Institute of Technology, Atlanta, GA, United States

We study optimal pricing strategies for service providers when customers may leave before receiving service due to long wait times. We allow service providers to set prices dynamically based on the current number of customers in the system. A customer joins the system if the quoted price is below their willingness to pay, which follows a known probability distribution. The service provider earns the quoted price as revenue when a customer joins the system but must fully reimburse any customer who abandons the system while waiting. The system incurs a per time per customer holding cost. We aim to determine the pricing policy that maximizes the long-run average profit. We first construct a continuous-time Markov chain with infinite capacity and prove that the system should reject new customers once the number of customers in the system exceeds a certain threshold. Using this result, we formulate the problem as a finite state Markov decision process. Our analysis reveals that the optimal price increases with the number of customers in the system, allowing us to use monotone policy iteration to solve for the optimal policy. We then prove that the optimal long-run average gain is monotone increasing in the service rate and the number of servers. Furthermore, the numerical experiments show our policy is robust to small parameter estimation errors. We provide guidelines for policy modification when parameters significantly deviate from their true values.

TC21

Building A Level 4 A407

Vehicle Routing and Fleet Management

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Michel Gendreau, Polytechnique Montreal, Montreal, QC, Canada

1 - Visits to the TSP-Drone Polytope : Bi-Cycle Elimination Inequalities and Two Alternative Compact Formulations

Taekang Hwang, KAIST, Daejeon, Korea, Republic of, Changhyun Kwon

In this study, we introduce two novel compact formulations, which are the Hamiltonian formulation and the Arc-ring formulation of the TSP-D problem. And we suggest facet-defining inequalities for the TSP-D polytope, named Bi-cycle elimination inequalities. The TSP-D is one of the main research streams investigating how we can use drone support efficiently in last-mile deliveries. In the TSP-D problem, in order to serve all customers, a truck and a drone cooperatively travel all the customer nodes without any repetition. Exploiting the cooperation of the truck and the drone, we can take advantage of using both types of vehicles at the same time which may lead to a shorter time to serve all customers. Despite the advantage, there has been almost no research on TSP-D polytopes, making it hard to invent powerful algorithms that can be used for practical purposes. For those reasons, we investigate TSP-D polytopes by introducing the two new compact formulations that we mentioned and studying their computational and theoretical properties. According to our investigation, compared to the existing TSP-D compact formulation, those are expected to be better formulations with regard to computational performances and their polyhedral properties. Furthermore, we demonstrate that we can obtain better lower bounds over small to medium-size instances when we add Bi-cycle elimination inequalities over the Hamiltonian formulation and, up to optimality, the inequalities tend to shorten the computational times over small-size instances. Due to the advantages, our understanding of the TSP-D polytope is expected to be helpful in building more powerful algorithms.

2 - A Logic-Based Benders Decomposition for the Integrated Ride-Sharing and Parking Allocation Problem

Hector Carlo, University of Puerto Rico - Mayaguez, Mayaguez, PR, Puerto Rico, Tomas Pérez-Romero

The Parking Allocation and Ride-Sharing System (PARS) is a non-pricing travel demand management strategy where drivers receive guaranteed parking in a venue if they are willing to participate in a centrally-coordinated carpool system. In PARS, riders request a ride from their depot to (and from) a particular venue and state their latest arrival and earliest departure times. Drivers request a parking reservation, including the car capacity, latest arrival time, and earliest departure time. The PARS problem is the optimization problem that seeks to design the feasible carpools to (and from) a venue such that a metric related to the carpools travel distance is minimizes and reserves parking. In PARS, individuals that are not selected as drivers are incorporated as riders and assigned to carpools.

A previously published efficient Mixed Integer Programming (MIP) formulation for the PARS problem is able to solve instances of medium size instances. This study proposes an innovative Logic-based Benders Decomposition algorithm (LBBD) to solve large instances of the problem. The LBBD strategy decomposes the PARS problem into a master problem (MP) and a feasibility subproblem (FS). The FS generates cuts (i.e., mathematical constraints) to guide the MP to a feasible solution, at which point it has reached the optimal solution. By the nature of the LBBD strategy, a lower bound is obtained after each iteration which is useful to test heuristics for very large instances of the problem. The proposed LBBD is able to solve PARS problem instances twice the size than the MIP using Gurobi.

3 - Probabilistic Models for Maximizing Service Area in Route Deviation Bus Transit Systems

Nikola Markovic, University of Utah, Salt Lake City, UT, United States, Madhu Mausam Thapa, Ye Chen, Shijie Chen, Yanshuo Sun, Ilya Ryzhov

This paper considers a route deviation system where buses can deviate from their fixed routes to meet on-demand requests within a specified buffer zone, thereby enhancing the efficiency and flexibility of conventional bus routes. The objective is to maximize the buffer zone width while accounting for desired headway, demand distribution, and the use of fixed stops along the route. Underused fixed stops provide opportunities for greater route deviations, while heavily used stops necessitate smaller deviations to maintain bus headway. We introduce both analytical probabilistic models and Bayesian simulation-based models to design and optimize this semi-flexible transit system. These models ensure that bus tour duration remains within the conventional transit schedule with high probability, accommodating on-demand requests submitted with appropriate notice. In the case study applied to the Utah Transit Authority's flex routes, significant improvements were observed. The results

demonstrate the system's ability to adapt to varying demand levels and operational conditions, providing a more responsive and passenger-centric public transportation service.

4 - Tactical Prepositioning of a Mixed Bus Fleet to Improve Real-Time Rescheduling in Adverse Conditions

Krystal Coleman, SUNY, University at Buffalo, Buffalo, NY, United States, Rajan Batta

Bus transportation systems are gradually replacing diesel and natural gas-fueled buses with battery electric buses (BEBs). During this process, fleets consist of a mix of vehicle types. The differences in strengths and weaknesses of BEBs and internal combustion engine buses (ICBs) create opportunities for operational resilience. We capitalize on these differences when deciding what composition of heterogeneous buses to assign to bus garages. Tactical prepositioning of the mixed fleet across bus garages may not significantly alter normal transit operations. However, in times of crisis, these decisions have a non-trivial impact on the system's resilience. This cascading effect justifies hierarchical decision making. We integrate vehicle allocation and bus scheduling as the upper-level task in a bi-level optimization problem. To minimize arrival delays or trip cancellations, the lower-level task involves rescheduling buses under adverse conditions or after disruption events. Joint optimization over a long-term planning horizon returns greater cost savings than addressing the bus scheduling and rescheduling problems separately. We use real data from NJ Transit and simulate the system's performance to extract insights into the passengers' experience. In simulations, we vary the assumptions that parametrize the optimization model and use statistical testing to validate whether those assumptions significantly impact the model's performance.

5 - A Variable Neighborhood Search Approach for the Electric Dial-A-Ride Problem

Michel Gendreau, Polytechnique Montreal, Montreal, QC, Canada

The goal of this research is to develop an effective solution approach for large Dial-A-Ride Problems with electrical vehicles (DARP-EV). This involves satisfying a set of transportation requests for passengers traveling between given pick-up nodes and drop-off nodes by transporting them using a fleet of electric vehicles (EVs). The vehicles are heterogeneous in terms of seating capacity but homogeneous in terms of battery capacity. Due to their limited battery capacities, EVs may need to recharge at charging stations during their routes. Charging operations can occur at any charging station when vehicles are empty, and partial recharging is allowed. Our solution method is based on the variable neighborhood search (VNS) metaheuristic. Starting from an initial solution s created using a linear programming-based heuristic, followed by a procedure for inserting charging stations, a neighborhood class is used in each iteration to generate a random solution to which a local search algorithm is applied. If the resulting solution is better, it replaces s , and we go back to the first neighborhood; if it is worse, s is not replaced, and we move to the next neighborhood. This whole process is repeated until a stopping criterion is satisfied. Seven different neighborhood classes (shakers) are used in the algorithm. We consider different cases with respect to the objection function being optimized and the charging functions used (linear vs concave piecewise linear). Computational experiments are performed on several sets of benchmark instances from the literature. We solve instances with up to 900 vehicles and 10,000 requests.

TC22

Building A Level 4 A408

Learning and Sequential Decision-Making

Invited Session

Applied Probability Society

Chair: Yilun Chen, CUHK Shenzhen, Shenzhen, N/A

1 - Statistical Inference for Markov Chains with Known Structure

Lin Fan, Northwestern University, Kellogg School of Management, Evanston, IL, United States, Peter Glynn

For Markov chains arising in many applications, the structure of the stochastic recursion defining the chain is known except for the distribution of the iid "driving noise" variables (the source of randomness in the chain). Given a dataset of such driving noise samples, and for a nonparametric model of the driving noise distribution, we study statistically efficient estimation of equilibrium and non-equilibrium expectation performance measures for the (general state space) Markov chain. Such expectations can be viewed as statistical functionals of the driving noise distribution, and we establish a framework based on semiparametric efficiency to characterize asymptotically minimum variance estimation for such functionals. In particular, we provide Foster-Lyapunov drift conditions to ensure suitable directional differentiability of the functionals and the existence of influence functions. Surprisingly, the nonparametric plug-in estimator, in which the empirical driving noise distribution is "plugged in" to the functional, can be inconsistent when the state space is infinite. To resolve the inconsistency issue, we introduce uniform versions of the classical Foster-Lyapunov drift conditions, which ensures that the conditions hold with the population kernels replaced by their empirical counterparts. Under such conditions, we establish asymptotic normality and efficiency of plug-in estimators of the performance measures of interest.

2 - Finite-Time Information-Theoretic Bounds in Queueing Control

Yunbei Xu, National University of Singapore, Singapore, Singapore, Yujie Liu, Vincent Tan

We establish the first finite-time information-theoretic lower bounds—and derive new policies that achieve them—for the total queue length in scheduling problems over stochastic processing networks with both adversarial and stochastic arrivals. Prior analyses of MaxWeight guarantee only stability and asymptotic optimality in heavy traffic; we prove that, at finite horizons, MaxWeight can incur strictly larger backlog by problem-dependent factors which we identify. Our main innovations are 1) a minimax framework that pinpoints the precise problem parameters governing any policy's finite-time performance; 2) an information-theoretic lower bound on total queue length; 3) fundamental limitation of MaxWeight that it is suboptimal in finite time; and 4) a new scheduling rule that minimizes the full Lyapunov drift—including its second-order term—thereby matching the lower bound under certain conditions, up to universal constants. These findings reveal a fundamental limitation on "drift-only" methods and points the way toward principled, non-asymptotic optimality in queueing control.

3 - Personalized Policy Learning Through Discrete Experimentation: Theory and Empirical Evidence

Ruohan Zhan, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, Zhiqi Zhang, Zhiyu Zeng, Dennis Zhang

Randomized Controlled Trials (RCTs), or A/B testing, have become the gold standard for optimizing various operational policies on online platforms. However, RCTs on these platforms typically cover a limited number of discrete treatment levels, while the platforms increasingly face complex operational challenges involving optimizing continuous variables, such as pricing and incentive programs. The current industry practice involves discretizing these continuous decision variables into several treatment levels and selecting the optimal discrete treatment level. This approach, however, often leads to suboptimal decisions as it cannot accurately extrapolate performance for untested treatment levels and fails to account for complex interactions between treatment effects and user characteristics. This study addresses these limitations by developing a theoretically solid and empirically verified framework to learn personalized continuous policies based on high-dimensional user characteristics, using observations from an RCT with only a discrete set of treatment levels. Specifically, we introduce a deep learning for policy targeting (DLPT) framework that includes both personalized policy value estimation and personalized policy learning. We prove that our policy value estimators are asymptotically unbiased and consistent, and the learned policy achieves a square root of n regret bound. We empirically validate our methods in collaboration with a leading social media platform to optimize incentive levels for content creation. Results demonstrate that our DLPT framework significantly outperforms existing benchmarks, achieving substantial improvements in both evaluating the value of policies for each user group and identifying the optimal personalized policy.

4 - When Learning Hurts: the Cost of Algorithmic Instability on Platforms

Jiaqi Lu, the Chinese University of Hong Kong (Shenzhen), Shenzhen, China, People's Republic of, Yilun Chen

When a platform uses conventional learning algorithms to allocate online demand to arms who are themselves agents, there can be an unintended consequence of highly fluctuating allocations in certain scenarios. This algorithmic instability raises concerns regarding the arms' welfare considerations such as fairness and risk-aversion. Motivated by this, we study two operational objectives for designing algorithms in stochastic multi-armed bandit environments, the expected regret and the standard deviation of the number of arm pulls. We characterize an interesting Pareto frontier: the product of the worst-case regret and the worst-case standard deviation of the number of arm pulls must be $\Omega(T^{\frac{3}{2}})$, asymptotically as the horizon length $T \rightarrow \infty$. This reveals a key impossibility result: any algorithm that is minimax optimal in expected regret must suffer from a standard deviation (in the number of pulls) of $\Theta(T)$ in the worst case; and any algorithm

having a standard deviation of $\sqrt{\Theta(T)}$ must incur a linear expected regret in the worst case. We then propose a simple class of UCB-f policies that is a natural generalization of the classic UCB1 policy, and show that it can achieve any point on the Pareto frontier by choosing the appropriate exploration function. We discuss the results' implications on the arms' welfare such as fairness and risk-aversion.

5 - Decision Making in Changing Environments: Robustness, Query-Based Learning, and Differential Privacy

Fan Chen, MIT, Cambridge, MA, United States

We study the problem of interactive decision making in which the underlying environment changes over time subject to given constraints. We propose a framework, which we call $\text{Hybrid Decision Making with Structured Observations}$ (hybrid DMSO), that provides an interpolation between the stochastic and adversarial settings of decision making. Within this framework, we can analyze local differentially private (LDP) decision making, query-based learning (in particular, SQ learning), and robust and smooth decision making under the same umbrella, deriving upper and lower bounds based on variants of the Decision-Estimation Coefficient (DEC). We further establish strong connections between the DEC's behavior, the SQ dimension, local minimax complexity, learnability, and joint differential privacy.

TC23

Building A Level 4 A409

Analytics for Simulation

Invited Session

Simulation

Chair: David Eckman, Texas A&M University, College Station, TX, United States

Co-Chair: Mohammadmahdi Ghasemloo, Texas A&M University, College Station, TX, United States

1 - Stochastic Sketchrefine: Decision-Making Under Uncertainty over Millions of Decision Variables

Riddho Haque, University of Massachusetts Amherst, Amherst, MA, United States, Anh L. Mai, Matteo Brucato, Azza Abouzied, Peter J. Haas, Alexandra Meliou

Decision-making under uncertainty often corresponds to choosing multisets ("packages") of probabilistic data tuples that collectively optimize expected outcomes while limiting risks. For large stochastic database relations, this involves solving integer programs having millions of decision variables. Monte Carlo methods can simulate numerous scenarios, i.e., sample realizations of the stochastic attributes of all the tuples, allowing for sample-average approximation (SAA). However, the number of scenarios needed for accurate approximation - and hence the size of the approximating deterministic optimization problem when using traditional SAA - increases with variance in the data, and the search space of the optimization problem increases exponentially with the number of tuples. Existing solvers take hours to process these problems on large relations containing stochastic attributes with high variance, often crashing before finding feasible results. We make two fundamental contributions towards scalably finding risk-constrained near-optimal packages. First, to handle high variance, we propose risk-constraint linearization (RCL), which converts a risk-constrained decision-making problem into an Integer Linear Program (ILP) whose size is independent of the number of scenarios used. Solving the ILP gives us a feasible and near-optimal package having an optimality guarantee. Second, to handle large relations, we propose Stochastic SketchRefine, a divide-and-conquer framework that breaks down a large stochastic optimization problem into subproblems involving smaller subsets of tuples, resulting in a final package with an overall optimality guarantee. Our experiments show that, together, RCL and Stochastic SketchRefine produce near-optimal results in orders of magnitude lower runtime than the state of the art.

2 - Global Sensitivity Analysis for Stochastic Models with Temporal Dynamics

Xi Chen, Virginia Tech, Blacksburg, VA, United States, Jingtao Zhang

Global sensitivity analysis is a critical tool for quantifying the influence of input parameters in complex models. Classical Sobol' indices, though widely used, have been primarily developed for deterministic computational models with a single output, limiting their applicability to stochastic systems with time-dependent, multidimensional outputs. In this work, we extend Sobol' indices to such settings by introducing a generalized Sobol' index that accounts for temporal dynamics. We propose three estimators for this index: a nested pick-freeze (NPF) estimator, a pick-freeze (PF) estimator, and a partition-based (PAR) estimator. Theoretical analysis shows that the NPF estimator achieves a mean squared error convergence rate of $O(B^{-2/3})$, while the PF and PAR estimators achieve a faster rate of $O(B^{-1})$, where B denotes the total number of model evaluations. Notably, the PAR estimator does not require a specific experimental design, enhancing its practical utility. We validate our approach through numerical experiments. The results reveal key temporal dynamics in model responses and highlight the advantages of the generalized Sobol' indices over classical indices in dynamic, stochastic settings.

3 - Automatic Generation of Models and Processes from Data

Zach Eyde, Arizona State University, Chandler, AZ, United States

Increasing demand for semiconductors from AI, HPC and automotive applications is driving the need for increased efficiency from manufacturers to keep pace. One major area of interest in both industry and academia is leverage digital twins (DT) to improve the efficiency of the semiconductor manufacturing process. An obstacle to the adoption of DTs in manufacturing is the development and maintenance of simulation and process models that power virtual representation. To tackle this, our research is focused on the automatic generation of modules and processes from production data, especially event logs from the equipment. We will demonstrate the full development flow from pre-processing the data to creating of graphs to represent the manufacturing process and finally incorporating the graphs into a simulation model. For our work, we preliminary look at queue networks to show the development steps and expand out to the manufacturing process of Assembly and Testing (A/T) of semiconductor devices.

4 - Moment-Based Approximations for Robust Optimization of Queueing Systems

Russell Barton, Penn State, University Park, PA, United States, Hyungkhee Eun, Sara Shashaani

Distributionally Robust Optimization (DRO) based on Wasserstein distance is preferred for general DRO settings where the computation of performance given the distribution form is straightforward. For complex queueing systems, the objective function evaluation requires computationally expensive simulation. We seek an effective proxy to identify input distributions with extreme system waiting times. Input distributions with large Wasserstein distance (from the nominal distributions) do not necessarily provide extreme objective values. We explore approximations for queue waiting time in a number of settings and show how they can provide low-cost guidance on extreme objective values, allowing for more computationally efficient DRO.

5 - Quantifying Uncertainty from Machine Learning Surrogate Models Embedded in Simulation Models

Mohammadmahdi Ghasemloo, Texas A&M University, College Station, TX, United States, David Eckman, Yaxian Li

Modern simulation models increasingly incorporate complex logic to emulate tactical decision-making supported by advanced decision support systems (DSSs). For practical reasons—such as computational burden, data privacy, or latency—modelers may replace embedded DSSs with machine learning (ML) surrogate models. We investigate the implications of using such surrogates on the estimation of system performance metrics. We distinguish this so-called surrogate uncertainty from conventional input uncertainty and propose new methods to quantify its impact. Specifically, we develop approaches based on bootstrapping and Bayesian model averaging to construct quantile-based confidence intervals for performance measures under surrogate approximation. We also introduce the use of regression-tree importance scores to apportion the total uncertainty between input variability and surrogate model error. Our methodology is demonstrated through a contact center simulation case study.

TC24

Building A Level 4 A410

The Low-rank Hypothesis in Modern Machine Learning

Invited Session

Applied Probability Society

Chair: Devavrat Shah, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Yassir Jedra, MIT, Cambridge, MA, United States

1 - Improving Low-Rank Adaptation of LLMs via Infinite-Width Analysis

Soufiane Hayou, UC Berkeley, Berkeley, CA, United States

Low-Rank Adaptation (LoRA) is a widely popular finetuning method for Large Language Models (LLMs). In this talk, I will show how we can leverage an asymptotic analysis of feature learning with LoRA in order to understand and improve initialization, learning rates, and target modules for LoRA finetuning. I will discuss the infinite-width dynamics of LoRA training via an operator called "Gamma" that tracks the impact of various hyperparameters on the infinite-width dynamics of LoRA.

2 - Near-Optimal Rank Adaptive Inference of High Dimensional Matrices

Frédéric Zheng, KTH, Stockholm, Sweden, Yassir Jedra, Alexandre Proutiere

We address the problem of estimating a high-dimensional matrix from linear measurements, with a focus on designing optimal rank-adaptive algorithms. These algorithms infer the matrix by estimating its singular values and the corresponding singular vectors up to an effective rank, adaptively determined based on the data. We establish, for the first time, instance-specific lower bounds for the sample complexity of such algorithms, uncovering fundamental trade-offs in selecting the effective rank: balancing the precision of estimating a subset of singular values against the approximation cost incurred for the remaining ones. Our analysis identifies how the optimal effective rank depends on the matrix being estimated, the sample size, and the noise level. We propose an algorithm that combines a Least-Squares estimator with a universal singular value thresholding procedure. We provide finite-sample error bounds for this algorithm. These bounds are better than the performance guarantees of existing rank-adaptive algorithms and they nearly match the derived fundamental limits. Our results rely on an enhanced analysis of matrix denoising methods based on singular value thresholding.

We apply our findings to Hankel matrix estimation in system identification by introducing Thresholded Ho-Kalman. Our algorithm further recovers both the system order and its Markov parameters, and we provide upper bounds for the sample complexity required to identify the system order and finite-time error bounds for estimating the Markov parameters. Interestingly, these bounds match those achieved by state-of-the-art algorithms that assume prior knowledge of the system order.

TC25

Building A Level 4 A411

Innovations in Last-Mile Delivery

Invited Session

TSL: Freight Transportation

Chair: Sara Reed, University of Kansas, Lawrence, KS, United States

1 - Last-Mile Delivery Routing with Parking Availability

Farima Salamian, University of Iowa, Iowa city, IA, United States, Ann Campbell, Sara Reed

Last-mile delivery is one of the most expensive parts of the supply chain. Thus, companies try to optimize the productivity of the delivery person via careful route planning. While route planning occurs prior to the driver starting the route, parking availability changes over time and can create implementation challenges for these route plans. However, with smart city technology, such as smart parking meters, companies may be able to incorporate parking availability information into their route planning to provide robust solutions. Unlike other models in the literature, we utilize this parking availability information to assist in route-planning. We formalize the routing problem as a Mixed Integer Program. As a baseline, we fix the route as a solution to the Traveling Salesperson Problem (TSP) based on driving times but must determine which parking spots to utilize and which customers to serve from each parking spot. For comparison, we then change the MIP to allow for optimizing the order of customers. We evaluate the performance over different levels of parking availability and numbers of customers. This allows us to (i) identify scenarios where the TSP order is not sufficient and (ii) provide routing techniques that incorporate parking availability information for last-mile delivery. These research questions are helpful for delivery providers to help improve their profitability, as well as for cities to understand the tradeoffs of investing in more sophisticated parking meter services in terms of reduction in congestion and other safety issues.

2 - Same-Day Delivery Integrated with Product Pickups

Mathias Klapp, PUC Chile, Santiago, Chile, Matias Gaete, Homero Larrain, Felipe Lagos

Same-day delivery (SDD) involves routing vehicles to deliver products from a warehouse to customers on the day orders are placed. Due to tighter deadlines and dynamic demand, SDD offers fewer consolidation opportunities than traditional delivery highlighting the need for alternative strategies to improve fleet utilization. We study an SDD operation integrated with product pickup requests, which represent customer returns or warehouse replenishments from local vendors. When first-mile (pickups) and last-mile (deliveries) are planned independently, requests in the same area are often handled by different vehicles. Integrating both operations can increase request density and improve operational efficiency. We aim to maximize the expected benefit of integrating the operation of both request types, each weighed differently. To this end, we design dynamic policies that adapt to the current fleet state and set of open requests while anticipating future arrivals, using a Multiple Scenario Approach combined with an ALNS heuristic for route planning. We estimate the value of integrated operations compared to a split-fleet alternative in computational experiments across varying requests, arrival rates, fleet sizes, and benefit weights. Results show that integration can increase total benefit by 8%. This improvement is more significant when request arrival rates are lower, the fleet is larger, or the benefit of serving deliveries relative to pickups is smaller.

3 - Recipient Preferences-Based Assignment of Parcels to Service Points

Margaretha Gansterer, University of Klagenfurt, Klagenfurt, Austria

The recent surge in e-commerce requires delivery companies to adapt and economize their logistic processes. Many parcels nowadays are delivered via automated service points (SPs), which prove to be an efficient, flexible, and sustainable alternative to home deliveries. Our study proposes a novel operation policy for parcel delivery firms. It is based on providing information about the availability of SPs in the recipient's vicinity and asking them to select their desired pickup location. Different recipients may have different sensitivities to delays and locations, leading them to make different selections. In contrast, the common practice in the industry is either to select the delivery SP solely based on the address of the recipient or based on the selection they made before real time information is known. We evaluate our preference-based policies against benchmarks that mimic common practice. We model a disutility function that represents the recipients' inconvenience due to the need to travel to the pickup locations and delays in the delivery process. We construct a simulation environment attempting to mirror the processes of parcel delivery for a realistic parcel delivery network. Our findings demonstrate the effectiveness of our preference-based policy, mainly when the recipient location preference variability is high or in moderately congested scenarios. We can significantly increase the recipient's satisfaction level and improve the load balance between the SPs in the network. Importantly, our policy is not only effective but also easy to implement, offering a promising avenue for enhancing service quality without the need for additional physical resources.

4 - Should Commercial Vehicles Double Park? Optimal Parking Decisions for Last-Mile Delivery in Urban Environments

Sara Reed, University of Kansas School of Business, Lawrence, KS, United States, Ann Campbell, Barrett Thomas

Parking is a challenge for last-mile commercial delivery drivers in urban environments. This work provides an analytical basis for understanding optimal parking decisions for commercial delivery drivers. We introduce the Stochastic Parking Problem (SPP) that determines whether a delivery driver should park at an available parking spot as well as whether the driver should double park at an unavailable parking spot at the risk of receiving a fine. We characterize the structure of the optimal policy for the SPP based on the probability that each parking spot is available and the expected cost of double parking. In particular, we show that the optimal policy for the SPP takes on one of four forms. We provide analytical results that determine which of the four forms achieves the optimal policy for a given problem instance of the SPP. Further, by exploiting the identified structure in these forms, we provide a polynomial time algorithm to determine the parameters of the optimal policy. Insights from this work

benefit both companies aiming to improve productivity of last-mile operations as well as cities aiming to control congestion, particularly congestion caused by double parking and cruising of commercial vehicles. We provide sufficient enforcement levels and fine amounts to eliminate double parking as well as improve utilization of limited parking infrastructure in urban environments.

TC26

Building A Level 4 A412

Bayesian Optimization

Invited Session

Simulation

Chair: Raul Astudillo, California Institute of Technology, Pasadena, CA, United States

Co-Chair: Peter Frazier, Cornell University, Ithaca, NY, 14850, United States

1 - Bayesian Optimization with Inexact Acquisition: is Random Grid Search Sufficient?

Chong Liu, State University of New York at Albany, Albany, NY, United States, Hwanwoo Kim, Yuxin Chen

Bayesian Optimization (BO) is a widely used iterative algorithm for optimizing black-box functions. Each iteration requires maximizing an acquisition function, such as the Upper Confidence Bound (UCB) or a sample path from the Gaussian process (GP) posterior, as in Thompson Sampling (TS). However, finding an exact solution to these maximization problems is often intractable and computationally expensive. Reflecting such realistic situations, in this work, we delve into the effect of inexact maximizers of the acquisition functions. Defining a measure of inaccuracy in acquisition solutions, we establish cumulative regret bounds for both GP-UCB and GP-TS without requiring exact solutions of acquisition function maximization. Our results show that under appropriate conditions on accumulated inaccuracy, inexact BO algorithms can still achieve sublinear cumulative regret. Motivated from such findings, we provide both theoretical justification and numerical validation for random grid search as an effective and computationally efficient acquisition function solver.

2 - A New Design Tool for Bayesian Optimization: the Gittins Index

Ziv Scully, Cornell University, Ithaca, NY, United States, Qian Xie, Theo Brown, Linda Cai, Raul Astudillo, Peter Frazier, Alexander Terenin

Many problems in practice do not fit cleanly into the classical Bayesian optimization setup, but are better modeled as Bayesian optimization "with a twist", or perhaps several twists. These twists include heterogeneous evaluation costs, multiple levels of fidelity, and partial feedback during evaluation. Handling any one of these twists alone, let alone a combination of them, is a significant challenge, and existing design principles (e.g. confidence bounds, expected improvement, Thompson sampling, knowledge gradient) do not always provide sufficient guidance about how to incorporate new twists.

We present a new design principle, based on the *Gittins index* from the theory of Markovian multi-armed bandits, that yields new policies for Bayesian optimization with any combination of these twists. The Gittins index is an acquisition function that yields the Bayes-optimal solution for a simplified version of Bayesian optimization where the objective function's domain is discrete, and its values at all points are mutually independent. Importantly, the Gittins index solves this simplified Bayesian optimization "with twists", thus guiding us towards Gittins-index-based policies for general Bayesian optimization "with twists". We demonstrate empirically that the Gittins index matches or outperforms state-of-the-art methods on realistic benchmarks. Among the obstacles we overcome is developing the first efficient algorithm that computes the Gittins index of a class of continuous-state Markov chains, namely linear Gaussian state space models.

3 - Newton-Step-Targeted Local Bayesian Optimization for High-Dimensional Black-Box Problems

Joel Paulson, The Ohio State University, Columbus, OH, United States, Wei-Ting Tang

Local optimization offers a compelling strategy for tackling expensive, high-dimensional black-box problems by bypassing the need to globally explore the entire domain. While traditional local methods often estimate descent directions using finite differences or random perturbations, these strategies can be sample-inefficient and produce noisy estimates – which is a particular challenge when derivatives cannot be directly observed. Local variants of Bayesian optimization (BO) provide an appealing alternative by building a probabilistic surrogate model of the objective and actively selecting informative samples to guide local search. Recent approaches, such as gradient-informed Bayesian optimization (GIBO), leverage this idea by reducing uncertainty in the estimated gradient and stepping in the direction of steepest descent. However, such first-order methods overlook curvature information, which plays a central role in accelerating convergence. In this work, we introduce NeST-BO, a new local BO framework that explicitly targets the Newton step, using both the gradient and Hessian estimates derived analytically from the Gaussian process surrogate model. Specifically, we develop a tractable acquisition function designed to reduce an upper bound on the one-step lookahead error in the Newton step and provide theoretical insight into its convergence properties. We demonstrate that NeST-BO achieves substantial gains over several existing local and global BO approaches across a range of challenging synthetic and real-world problems (ranging from 10 to 50 dimensions). In all cases, NeST-BO achieves faster convergence and lower regret than existing methods, establishing it as a powerful tool for efficient black-box optimization in high-dimensional settings.

4 - Scalable Bayesian Optimization of Composite Functions for High-Dimensional Inverse Problems in Electron Microscopy

Poompol Buathong, Cornell University, Ithaca, NY, United States, Dasol Yoon, Chia-Hao Lee, David Muller, Peter Frazier

We consider inverse problems from medical imaging and electron microscopy where we seek to recover a vector of latent parameters x by minimizing the recovery error $\|h(x) - y\|^2$. Here, y is the observed image, expressed as a vector, and $h(x)$ is a time-consuming vector-valued black-box derivative-free simulator. When y is low-dimensional, Bayesian Optimization of Composite Functions (BOCF), a previously-proposed method, effectively solves problems of this form. It models $h(x)$ with a multi-output Gaussian process (GP) and uses the fact that the objective function is a composition of $h(x)$ with a known simple function to reduce the number of costly evaluations of $h(x)$ needed for low recovery error. When y is high-dimensional, the computational time and memory requirements of high-dimensional multi-output GPs make this approach impractical. We propose a scalable alternative that groups components of $h(x)$ into lower-dimensional outputs, reducing the number of surrogates while continuing to support accurate query-efficient optimization. Demonstrated on an electron microscopy problem, where we determine the specimen thickness and object misorientation with Bayesian optimization. Our method successfully finds low-recovery-error solutions with fewer evaluations of $h(x)$ than standard Bayesian optimization and achieves significantly faster runtime than the estimated runtime of full BOCF.

5 - Marobo-Se: Multi Agent Rollout for Bayesian Optimization - Sample Efficient

Tanmay Khandait, Arizona State University, Tempe, AZ, United States, Giulia Pedrielli

Bayesian Optimization (BO) is a sample-efficient strategy for optimizing expensive black-box functions and has been widely adopted in various domains. However, BO's performance can degrade in high-dimensional or multi-modal settings, where optimizing the acquisition function becomes computationally challenging and prone to poor

local solutions. In this work, we introduce MAroBO-SE, a general-purpose, algorithm-agnostic wrapper that improves sampling efficiency by augmenting any existing BO method with coordinated multi-agent rollouts. MAroBO-SE deploys multiple agents that explore the acquisition landscape via stochastic trajectories, enhancing diversity and exploration without requiring changes to the surrogate model, acquisition function, or optimization routine. As a result, MAroBO-SE preserves the base BO algorithm's structure while improving robustness and performance in challenging optimization settings. We demonstrate that MAroBO-SE consistently improves optimization outcomes when integrated with state-of-the-art methods such as TuRBO, achieving superior results on both synthetic benchmarks and real-world tasks.

TC27

Building B Level 2 B201

Machine Learning for Optimization

Invited Session

Computing Society

Chair: Can Li, Purdue University, West Lafayette, IN, United States

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

1 - Scientific Machine Learning for Optimization and Control

Jan Drgona, Johns Hopkins University, Baltimore, MD, United States

This talk presents a scientific machine learning perspective (SciML) on modeling, optimization, and control. Specifically, we will discuss the opportunity to develop a unified SciML framework for modeling dynamical systems, learning to optimize, and learning to control methods. We demonstrate the application of these emerging SciML methods in a range of engineering case studies, including modeling of networked dynamical systems, building control, and dynamic economic dispatch problem in power systems. Furthermore, we will introduce the NeuroMANCE open-source library, facilitating the implementation and prototyping of diverse SciML methods for a broad range of application problems.

2 - Learning Hybrid Representations of Differential-Algebraic Systems Through Nonlinear Programming Approaches

Laurens Lueg, Carnegie Mellon University, Pittsburgh, PA, United States, Victor Alves, Carl Laird, Lorenz Biegler

We consider so-called neural differential-algebraic systems of equations (neural DAEs), where some unknown relationships within a DAE are learned from available data. Training neural DAEs, similarly to neural ordinary differential equations (neural ODEs), is computationally expensive, as it usually requires the solution of a DAE for every parameter update. Furthermore, the rigorous consideration of algebraic constraints is difficult within common deep learning training algorithms such as stochastic gradient descent (SGD), which are fundamentally designed for unconstrained optimization. Therefore, in this work, we apply the so-called simultaneous approach to neural DAE problems. This results in a fully discretized nonlinear optimization problem that is solved to local optimality and simultaneously obtains the neural network parameters and the solution to the corresponding DAE. This eliminates the need for costly sequential calls to ODE/DAE solvers. Our presented framework is of general nature, with explicit consideration of hybrid models, where some components of the DAE are known, e.g. physics-informed constraints. Furthermore, we present a systematic strategy for improving the performance and convergence of the nonlinear programming solver, through specialized initialization strategies and appropriate Hessian approximations. Lastly, we provide several promising future directions to improve the scalability and robustness of our approach.

3 - Learning to Optimize for Mixed-Integer Non-Linear Programming with Feasibility Guarantees

Bo Tang, University of Toronto, Toronto, ON, Canada, Elias Khalil, Jan Drgona

Mixed-integer nonlinear programs (MINLPs) arise in domains as diverse as energy systems and transportation, but are notoriously difficult to solve, particularly at scale. While learning-to-optimize (L2O) methods have been successful at continuous optimization, extending them to MINLPs is challenging due to integer constraints. To overcome this, we propose a novel L2O approach with two integer correction layers to ensure the integrality of the solution and a projection step to ensure the feasibility of the solution. We prove that the projection step converges, providing a theoretical guarantee for our method. Our experiments show that our methods efficiently solve MINLPs with up to tens of thousands of variables, providing high-quality solutions within milliseconds, even for problems where traditional solvers and heuristics fail. This is the first general L2O method for parametric MINLPs, finding solutions to some of the largest instances reported to date.

4 - Systematically Comparing Embedded Surrogates for Optimization

Ali Asger, Carnegie Mellon University, Pittsburgh, PA, United States, Bashar Ammari, Logan Matthews, R. Cory Allen, Dimitri Papageorgiou, Carl Laird

The optimization community has witnessed a marked increase in the utilization of surrogate models in the design, development, and/or operation of complex systems and enterprise-wide supply chains. Libraries like OMLT (Optimization and Machine Learning Toolkit) and Gurobi's Machine Learning package allow users to automatically embed trained machine learning models into optimization models. With numerous surrogate options now available, an open question is: Given trained surrogates of tantamount predictive accuracy on a test set, which surrogate model leads to the best optimization performance in terms of overall solution quality and solution time? After adapting well-known benchmark MILP instances, we compare a representative surrogate from each of the following classes: (1) ReLU-based neural networks; (2) standard decision trees; (3) difference-of-convex piecewise linear models; (4) linear model decision trees; (5) gradient boosted trees.

5 - Learning to Tune Low-Fidelity Surrogates for Graph-Structured Optimization with One Training Example

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

In this talk, we present a machine learning framework for training low-fidelity surrogate models tailored to optimization problems with underlying graph structures, using only a single high-fidelity training example. The goal is to enable efficient generalization to unseen graph instances by leveraging shared structural patterns across problems. The proposed framework consists of two key components: (i) a graph neural network that maps instance-specific features (e.g., topology, node or edge attributes) to the optimal tunable parameters of the low-fidelity model; (ii) a differentiable optimization layer that embeds the surrogate directly into the learning pipeline, allowing for end-to-end gradient-based training; and (iii) an implicit loss function that quantifies the mismatch between the low- and high-fidelity models, even when the high-fidelity model is non-differentiable. This architecture enables one-shot training of generalizable surrogates, without requiring labeled instances. We validate the framework on the AC optimal power flow problem using the DC-OPF as a tunable surrogate, demonstrating strong performance in terms of accuracy, feasibility, and inductive generalization on test systems with up to 13,659 buses.

TC28

Building B Level 2 B202

Frontier Applications of Machine Learning

Invited Session

Computing Society

Chair: Maura Hegarty, Massachusetts Institute of Technology, Cambridge, MA, 02141, United States

1 - A Unified View of Optimization in the US Transplant System

Maura Hegarty, Massachusetts Institute of Technology, Cambridge, MA, United States, Dimitris Bertsimas

The allocation and utilization of deceased-donor organs in the U.S. transplant system is governed by a fragmented, organ-specific decision process. Separate committees—Heart, Lung, Kidney, Pancreas, and Liver & Intestine—along with topic-focused groups (e.g., Living Donor, Histocompatibility) and an ad hoc Multi-Organ Committee, each prescribe distinct policies and allocation rules. While this vertical structure addresses organ-specific clinical nuances, it overlooks the reality that a single donor

frequently provides multiple transplantable organs that can be donated in pairs and triplets. As a result, current policy deliberations may inadvertently sacrifice overall patient benefit in favor of siloed decision-making. We propose a unified optimization framework that jointly considers all allocation and utilization decisions for a donor's full organ yield. Our approach models the transplant system as a large-scale, seeking to maximize aggregate patient-years of life and quality of life.

2 - Generative AI for Personalized Instruction

Romain Puech, MIT, Cambridge, MA, United States

One-to-one tutoring is one of the most efficient methods of teaching. With the growing popularity of Large Language Models (LLMs), there have been efforts to create LLM-based conversational tutors, which can expand the benefits of one-to-one tutoring to everyone. However, current LLMs are trained primarily to be helpful assistants and lack crucial pedagogical skills. For example, they often quickly reveal the solution to the student and fail to plan for a richer multi-turn pedagogical interaction. To use LLMs in pedagogical settings, they need to be steered to use effective teaching strategies: a problem we introduce as Pedagogical Steering. We develop StratL, an algorithm to optimize LLM prompts and steer it to follow a predefined multi-turn tutoring plan represented as a transition graph. We use this algorithm to implement an AI problem-solving tutor for MIT Open Learning.

3 - Interpretable AI-Driven Guidelines for Type 2 Diabetes Prevention and Treatment

Dewang Agarwal, Massachusetts Institute of Technology, Cambridge, MA, United States, Dimitris Bertsimas

Objective: To develop precise, structured, data-backed guidelines for type 2 diabetes prevention and treatment progression, suitable for clinical adoption.

Research Design and Methodology: For prevention, we analyzed data from two randomized controlled trials: The Diabetes Prevention Program/Outcomes Study (DPP/DPP-OS) and ACT-NOW, both of which followed high risk individuals under different treatments. We created counterfactual models for treatment specific survival curves, after which tree-based AI models identified the optimal intervention.

For treatment prescription, we used observational data from Boston Medical Center (BMC) from 1998-2014. Patient visits were stratified by pre-visit treatment and then recommend the optimal treatment. Using optimization and machine learning, we reduced confounding bias to emulate randomized trial conditions. Multiple tree based models were trained and manually integrated into end-to-end prescription pipelines, one for each patient group. Performance was evaluated on held-out BMC data and externally validated on a Hartford dataset (visits between 2020-2024).

Results: The prevention pipeline was predicted to reduce T2DM incidence by 10% in high-risk patients. The prescription pipeline achieved median reductions in HbA1c (blood sugar) of 0.26% and 0.13% above physician performance at BMC and Hartford respectively.

Conclusion: Interpretable AI-driven guidelines for T2DM prevention and treatment outperform standard care and are ready for real world clinical deployment to improve patient outcomes.

4 - A Machine Learning Approach to Fluid Markov Decision Processes

Lisa Everest, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA, United States, Dimitris Bertsimas, Cheol Woo Kim, Jose Nino Mora

We propose a machine learning approach to the optimal control of fluid Markov decision processes (MDPs) with state equations that are affine in the state variables. We derive fundamental properties of the MDP problems, which we then use to design a numerical algorithm. This algorithm begins with solving MDP instances with varying initial states to generate a dataset. We train an Optimal Classification Tree with hyperplane splits (OCT-H) to learn a control policy from this dataset.

5 - Adaptive Forests for Classification

Yubing Cui, MIT, Cambridge, MA, United States, Dimitris Bertsimas

We propose **Adaptive Forests (AF)**, a novel method that assigns input-dependent weights to CART trees using Optimal Predictive-Policy Trees and Mixed Integer Optimization. AF consistently outperforms Random Forests, XGBoost, and other weighted ensembles across 20+ datasets in both binary and multi-class classification tasks.

TC29

Building B Level 2 B203

Algorithmic Approaches to Access and Affordability in Power Systems Optimization

Invited Session

ENRE: Energy

Chair: Rabab Haider, University of Michigan, Ann Arbor, MI, United States

1 - Integration of Critical Public Safety Needs Within Grid Planning and Optimization Methods

Jesse Dugan, National Renewable Energy Laboratory, Golden, CO, United States, Jordan Burns, Sara Peterson, Maya Fein-Cole

Extreme weather events and the threat of cyberattacks, coupled with the growing interdependence of infrastructure systems, have made resilience a critical objective for energy systems. Power outages and infrastructure disruptions pose significant societal risks, threatening public safety, the economy, and national security. Investments in distribution system resilience through grid modernization and hardening can help mitigate the impacts of long-duration power outages. However, traditional models and metrics used to assess infrastructure resilience often focus solely on the performance of physical infrastructure rather than the performance of the societal systems supported by infrastructure. This work presents a methodology to quantify the impacts of power outages on public safety by analyzing their effects on the availability of and access to critical services. This approach integrates grid models with residential and critical service locations, enabling us to explore outage consequences in a detailed spatial and temporal context. We employ routing algorithms, service value assignment, and travel cost analysis to examine the ways in which power outages impact households by decreasing their access to critical services. By integrating this methodology into power systems planning and operational decisions, we assess the performance of grid hardening and modernization investments from the public safety perspective. An application of this methodology in South Seattle, WA demonstrates how it can be used to site distribution automation equipment that optimizes public safety by providing households with continuous access to essential services during outages.

2 - Metrics to Support High-Resolution Community Impact Considerations in Electric Power System Decisions

Abigail Beck, University of Houston, Houston, TX, United States

The provision or absence of electricity has collective implications across an entire community, yet impacts from service disruptions have been shown to not occur uniformly but lead to disparate impacts on certain vulnerable populations (e.g., elderly, individuals with electric dependent medical devices). These populations can also be serviced by infrastructure more prone to outages only to exacerbate the likelihood of outage impacts. To improve community resilience these impact differences must be considered in power system optimization decisions (e.g., retrofit, restoration, operation, etc.). Aggregate community-wide metrics have previously been defined to assess broad system-level inequalities. However, high-resolution metrics are required to better detect the potential differential outage impacts and to target actions to reduce impact burdens on different groups or areas of concern. This talk presents high-resolution Theil's T-based metrics which are flexible to consider any community impact measure of concern and be implemented within multiple power system optimization decisions. The Theil's T-based metrics measure both the overall inequality and the amount a certain individual or group contributes to the overall inequality in impact. These metrics can be implemented within any power system decision to identify

inequalities, produce prioritizations, and formulate an objective function. The metric and its assessment and prioritization capability are demonstrated in simplified decision cases for distribution pole retrofitting and restoration. These applications demonstrate the metric's ability for high resolution impact-based prioritization and lays the groundwork for future optimization-based implementations.

3 - Embedding Affordability Goals into Grid Planning via Differentiable Optimization

Samuel Talkington, Georgia Institute of Technology, Atlanta, GA, United States, Amanda West, Ryan Piansky, Rabab Haider, Daniel Molzahn

With rapid load growth, extreme weather, and high penetration of intermittent generation, maintaining affordable access to electricity is of critical concern for system operators and utilities. This talk introduces a computational framework, Locational Marginal Burden (LMB), that is grounded in differentiable optimization. The LMB framework emerges as an intrinsic property of optimal power flow (OPF) problems. By differentiating through the dual solutions of OPF problems, i.e. the wholesale electricity prices, we reveal the impact of power system operating principles and grid infrastructure on affordability. The LMB framework endogenizes the energy burden experienced by customers (a metric for energy affordability) directly into grid optimization models. In doing so, the LMB operationalizes energy burden, providing a way to incorporate affordability considerations into planning and operating models. We propose the LMB framework as a way to support policymaking and investment decisions, such as line upgrades and capacity expansion while meeting critical energy affordability goals. This approach provides a mathematically grounded pathway for embedding equity considerations in power systems optimization.

4 - Energy Management for Prepaid Electricity Customers

Maitreyee Marathe, University of Massachusetts Amherst, Amherst, MA, United States

For a large and recently increasing number of households, affordability is a major hurdle in accessing sufficient electricity and avoiding service disconnections. Such households need to actively ration the amount of energy they use by deciding which appliances to use and for how long. In this work, we consider prepaid customers – a particularly vulnerable group of customers who are required to pay for their electricity a-priori and get immediately disconnected if their prepaid wallet runs out of balance. Low-income customers and prepaid customers have been largely neglected in home energy management literature. We present a tool that helps such households extend the availability of their critical appliances by limiting the use of discretionary ones, and prevents disconnections. The tool is based on a linear optimization problem that only uses average power demand forecasts and can be solved to optimality using a simple greedy approach. Our model accounts for daily variations in the relative priority of appliances and time-of-use pricing. Furthermore, we demonstrate the computational simplicity of our model by implementing it on an inexpensive 8-bit microprocessor. In a numerical case study based on real household data, we compare the performance of our model with two mixed-integer linear programming models that require more detailed demand forecasts and optimization solvers for implementation. Our results show that our proposed linear model is much simpler to implement, while providing similar performance under realistic circumstances, lending itself well to adoption across different regions and utility programs.

5 - Reducing the Academia Discourse Divide: How to Rethink Collaboration on Environmental Justice and Equity

Aneesha Manocha, University of California Berkeley, Berkeley, CA, United States

Various disciplines (engineering, social sciences, and humanities) have historically approached questions of environmental justice (EJ) and equity differently, from how they formulate the scope and impacts of the problem to how they propose enacting change and recommendations. While these broad disciplines have contributed immensely to problems related to EJ, there are opportunities for deeper interdisciplinary collaboration, especially in planning for clean, just energy transitions. In this piece, we present an EJ framework, where we trace epistemologies of how various disciplines conceptualize knowledge to their methods and how they present and publish work. We extensively survey the EJ literature across academic disciplines to determine how different fields discuss similar issues and can learn from each other. We highlight opportunities to facilitate collaboration through a few case studies. We conclude with recommendations on how researchers can work with scholars across fields who think about data, methods, and research outputs differently.

TC30

Building B Level 2 B204

Distributed Energy Resource Integration: Statuses and Challenges

Invited Session

ENRE: Electricity

Chair: Feng Pan, Pacific Northwest National Laboratory, Richland, WA, United States

Co-Chair: Hugo Santarem de Araujo, Cornell University, Ithaca, NY, United States

1 - Price Consistency and Uplift in Wholesale Electricity Markets

Hugo Santarem de Araujo, Cornell University, Ithaca, NY, United States, Jacob Mays, Han Shu

Price formation is central to power market operations, but is complicated by two persistent challenges: non-convexities in unit commitment and the uncertainty introduced by variable renewable energy. While each issue has been studied independently, their combined effects remain poorly understood, especially in the sequential decision-making context. One consequence of sequential decisions is the occurrence of mismatches between provisional prices (based on initial schedules) and binding prices (those that ultimately clear the market). These mismatches often give rise to uplift payments, which keep generators financially whole when dispatch decisions deviate from profit-maximizing decisions. Due to their non-transparent and discriminatory nature, market operators and regulators have implemented price formation reforms seeking to limit uplift payments. Many of these reforms are inspired by Convex Hull Pricing (CHP), which has been shown to minimize uplift in deterministic settings. Our study examines how non-convexity and uncertainty jointly shape price formation. We model the unit commitment problem sequentially over time, reflecting how Independent System Operators (ISOs) operate in practice, and compare the performance of price formation alternatives. Simulations use a stylized model of the New York ISO grid, incorporating thermal, solar, wind, and hydro resources totaling approximately 40,000 MW under three different operational policies. For each run, we compute both provisional and binding prices and the corresponding uplift payments. Results show that while CHP reduces expected payments, it has little impact on actual binding payments. Instead, the choice of operational policy has a larger impact on price consistency and uplift.

2 - How to Procure Green Power for Real Emissions Reductions?

Yago del Barrio, Johns Hopkins University, Baltimore, MD, United States

This work investigates how corporate green power procurement strategies interact with national and regional carbon and renewable energy policies. It examines how procurement design choices, such as temporal and spatial matching requirements, affect system-wide emissions and cost outcomes. Multiple optimization frameworks are employed to reflect different assumptions and perspectives. In particular, a bilevel optimization model is developed to analyze the potential influence of large corporate buyers on market operations and emissions trajectories. This analysis is especially relevant given the rapid expansion of energy-intensive data centers, which are often concentrated in specific regions and can significantly impact local grid dynamics. The framework is applicable to other large procuring corporations seeking to ensure that their clean energy commitments lead to real emissions reductions.

TC31

Building B Level 2 B205

Resilience, Simulation, and Infrastructure Risk

Contributed Session

Contributed

Chair: Godwin Obi, Princeton University, 88 College Road West, Princeton, NJ, 08544, United States

1 - Multi-Island Power System Optimization: a Feasibility Study of the Eastern Caribbean Power Pool

Avery Barnett, Princeton University, Princeton, NJ, United States, Robert Brecha, Olivia Roslansky, Jesse Jenkins

Nearly all countries and territories in the Eastern Caribbean face high electricity costs and energy insecurity due to aging grids, small-scale demand, and above all, due to reliance on imported fossil fuels. To overcome this, we explore the feasibility of an Eastern Caribbean Power Pool (ECPP), a proposed interconnected grid that spans from Trinidad and Tobago to Antigua and Barbuda. This is the first study of its kind to apply a multi-island capacity expansion framework with geopolitical feasibility assessments for a Power Pool in the Eastern Caribbean. We use GenX, a capacity expansion model, to determine least cost generation and transmission investments and apply Modelling to Generate Alternatives (MGA) to identify a set of near-optimal interconnection pathways via subsea cables. This methodology seeks to balance technical feasibility, cost, and the resource potential within each country. Scenarios are evaluated for their ability to support the national renewable energy targets while maintaining or lowering electricity costs. We assess implementation feasibility by comparing model outcomes with qualitative geopolitical factors based on prior research. The results will offer insight to regional energy planning and provide a foundational analysis for the Caribbean Development Bank's CREGI-RES initiative.

2 - Data Injection Attacks on Distributed Systems: An Information-Theoretic Approach

Xiuzhen Ye, Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC, United States, Wentao Tang

The unprecedented data acquisition capabilities in modern cyber-physical systems (CPSs) have increased their vulnerability to malicious attacks. Typical CPSs—such as smart grids and chemical plants—are susceptible to threats such as data injection attacks (DIAs). As a fundamental requirement for ensuring safety and reliability, investigating DIAs in these domains is of critical importance. We are particularly interested in *distributed attack* and *sparse attack* strategies, which become feasible for attackers due to the integration of computing and communication components in CPSs.

This work considers DIA design from the attackers' standpoint, who are able to exploit the statistical properties of the system's state variables. To quantify the statistical information with fundamental metrics and thus effectively leverage them in optimal attack constructions, we adopt an *information-theoretic* approach. Specifically, the attackers aim for targeting the sensor measurements at the stationary operating points in the system, with the objective to disrupt the distribution of the state variables while remain stealthy. To this end, mutual information and Kullback-Leibler divergence are used to measure the attack disruption and stealthiness, respectively.

We hence use an optimization formulation under the restrictions of distributed attacks (where each attacker has access to an assigned set of sensors) and sparse attacks (where the number of sensors attacked should be limited). As a realistic solution, by modeling the interactions between attacks in different locations of distributed systems, *game-theoretic strategies* are proposed. Case studies on smart grids and chemical reactor systems are presented to illustrate the proposed method with comparison to *full attacks*.

3 - Port Resilience in Disruptive Scenarios: a Milp-Based Decision Support Model

Ruben Yie-Pinedo, Universidad el Norte, Barranquilla, Colombia, Johanna Padilla, Adriana Moros

Port resilience is essential for maintaining operational continuity and efficiency amid global supply chain disruptions. This study introduces a Mixed-Integer Linear Programming (MILP) model for the Discrete Berth Allocation and Quay Crane Assignment Problem (BACAP), enhanced by a quantitative resilience index. The model aims to optimize berth and crane scheduling in container terminals under systemic disruption scenarios.

Currently under development, the model explores six operational scenarios involving disruptions such as vessel delays, crane unavailability, and partial berth closures. A composite resilience index guides decision-making, built using Principal Component Analysis (PCA) and the Entropy Weight Method (EWM) on key performance indicators (KPIs). Bootstrap resampling ensures the statistical robustness of the indicator weights.

By integrating this resilience index into BACAP, the model enhances operational flexibility, reduces service time variability, and supports greenhouse gas mitigation. A case study in a Caribbean port will validate the model's applicability, aiming to improve port management and sustainability in uncertain environments.

This work offers a quantitative framework for resilient planning in port terminals, aligning operational efficiency with environmental sustainability and strengthening the competitiveness of maritime supply chains.

4 - Linear Ac Power Flow for Optimal Transmission Switching

Sergio A. Dorado-Rojas, Georgia Institute of Technology, Atlanta, GA, United States, Constance Crozier, Daniel Molzahn

Optimal transmission switching (OTS) is a valuable tool for enhancing grid operations, allowing system operators to achieve objectives like congestion reduction, cost minimization, and wildfire risk minimization. Conventionally, OTS problems are formulated with highly simplified (DC) or computationally intractable (AC) power flow models. The full AC optimal transmission switching (ACOTS) problem presents a significant computational challenge with its binary switching variables and nonconvex AC power flow constraints. While the DC optimal transmission switching (DCOTS) problem offers a tractable linear approximation, it sacrifices accuracy. This work introduces a linear AC power flow model into an OTS formulation, resulting in the linear ACOTS (LACOTS) problem. Despite its potential, this linear AC formulation is rarely used as a baseline in OTS. We demonstrate LACOTS's capability to serve as an effective linear surrogate for the nonlinear AC power equations through a combination of theoretical analysis and numerical experiments. Our ongoing efforts are concentrated on systematically comparing DCOTS and LACOTS solutions across a range of test cases, aiming to fully characterize the benefits and illuminate the power of this often-overlooked linear AC approach.

5 - Securing Firm Gas Capacity: Impacts on System Costs and Gas Plant Competitiveness of on-Site Fuel Storage in Current and Future U.S. Power Grids

Godwin Obi, ZERO Lab, Princeton University, Princeton, NJ, United States, Jesse Jenkins

Fuel supply disruptions during extreme weather events—such as Winter Storm Uri, which caused hundreds of billions in economic damage—highlight the vulnerability of gas-fired power plants and the importance of firming fuel supply. On the other hand, as renewables scale and gas plants face lower utilization rates, traditional financial mechanisms for pipeline-based fuel delivery may become harder to justify. This study examines the system-wide impacts of requiring 1 to 14 days (2 weeks) of on-site fuel storage, focusing on cost implications and deployed gas capacity in the short term (under current IRA policies and a no-IRA-policy scenario) and long term (under high-renewables and business-as-usual scenarios) across the U.S. grid.

Using GenX, a state-of-the-art capacity expansion model, we quantify changes in system fixed and fuel costs and shifts in gas capacity with increasing storage duration. We also conduct a breakeven cost analysis to estimate, under both policy scenarios, the capital costs at which LDES and SMRs could displace gas capacity with plausible fuel storage. We evaluate three storage options—diesel, LNG, and liquid hydrogen—with a focus on: What storage medium is most plausible? How do system and plant-level costs change? At what storage duration or capital cost do clean firm alternatives replace gas?

Results show 1–14 days of storage raise system costs by up to 3%, plant fixed costs by over 6%, and reduce gas capacity by nearly 2%, with diesel as the most cost-effective option. At 2 weeks, diesel storage needs reach one-quarter of U.S. annual commercial consumption.

TC32

Building B Level 2 B206

Open-Source Tools for Modeling and Simulation in Power Systems Research

Invited Session

ENRE: Energy-Climate

Chair: Feng Qiu, Argonne National Laboratory, Lemont, IL, United States

1 - Unitcommitment.Jl: A Julia/Jump Optimization Package for Security-Constrained Unit Commitment

Alinson Santos Xavier, Argonne National Laboratory, Lemont, IL, United States

Security-Constrained Unit Commitment (SCUC) is one of the most fundamental and challenging problems in power systems optimization, being solved daily by Independent System Operators (ISOs) to clear the electricity markets. In this talk, we present UnitCommitment.jl, an open-source optimization package designed to help researchers overcome common obstacles in developing and evaluating new solution methods for this problem. The package offers several features, including: (i) a flexible and well-documented data format for the problem, created with input from ISOs, that makes it easier for researchers to share data sets across institutions; (ii) a collection of large-scale benchmark problems, gathered from the literature and extended with data-driven methods to make them more complex and realistic; and (iii) implementations of state-of-the-art mathematical formulations and solution methods for the problem; and (iv) automated benchmark scripts to evaluate the performance of newly proposed methods accurately.

2 - Open-Source Tools for Modeling and Simulation in Power Systems Research

Wei Gao, Argonne National Lab, Lemont, IL, United States, Feng Qiu, Shijia Zhao

Reliability analysis is critical for ensuring secure and continuous power system operation under both normal and stressed conditions. With increasing grid complexity driven by renewable integration and extreme events, there is a growing demand for computational tools that are robust, scalable, and fast. Traditional numerical methods, such as Newton-Raphson iterations, often suffer from convergence issues and inefficiencies, particularly in large-scale contingency analysis.

This work presents PowerSAS.jl, a high-performance, Julia-based framework for power system reliability analysis built on the semi-analytical solution (SAS) methodology. SAS approximates system dynamics using power series expansions, yielding closed-form expressions over time intervals. This enables faster simulations, larger time steps, and improved numerical stability compared to iterative methods.

PowerSAS.jl supports steady-state, transient, and extended-term simulations, including N-1 contingency screening, voltage and frequency stability assessment, and cascading failure analysis. It inherits key capabilities from SAS: high numerical robustness, support for differential-algebraic systems, and the ability to simulate power systems with over 200,000 buses. Enhanced with Julia's parallel computing strengths, PowerSAS.jl achieves significant speedup and scalability improvements.

Released as open-source, PowerSAS.jl provides a flexible and transparent tool for planners and researchers, accelerating comprehensive reliability assessments essential for real-time and market operations.

3 - Andes Journey: Ten Years of Open-Source Innovation in Power Systems

Hantao Cui, North Carolina State University, Raleigh, NC, United States

ANDES, a Python software for power system simulation, started as a weekend project for learning Python and power system fundamentals. Ten years later, it is a full-fledged open-source simulator with tens of thousands of lines of code that is widely used for modeling, simulation and analysis. This talk reflects on the journey for getting there on both the technical decisions and the human side. Primarily, it took two major rewrites and deliberate efforts of community building to keep the project relevant and easy to use for both students and professionals. The talk will also discuss the sustainability of opensource software, focusing on tools in the engineering domains that typically attract a small pool of contributors. In addition, I will share the vision of an opensource ecosystem where a spectrum of opensource tools can develop synergy for applications such as production optimization, stability simulation, data-drive decision making, and climate impact studies.

4 - Diff2sp: Diffusion Models for Correlated Scenario Generation in Stochastic Programming

Haixiang Sun, Purdue, West Lafayette, IN, United States, Andrew Liu

The quality of scenario generation has a direct impact on the performance of decisions made under stochastic programming frameworks. Traditional Monte Carlo and supervised learning approaches are often computationally intensive, require large sample sizes or labeled data, and struggle to capture complex temporal and cross-variable dependencies. In this work, we introduce Diff2SP, a diffusion-based framework that learns joint uncertainty distributions from historical observations by embedding the downstream optimization objective into its training loss, so that the generator learns to prioritize uncertainty patterns that most impact decision outcomes.

We derive a finite sample regret bound that relates the Wasserstein distance between the learned and true distributions to loss in objective value and show that the bound achieved by Diff2SP is strictly tighter than the best available guarantee for Generative Adversarial Networks (GANs). Extensive evaluation supports the theory. In our numerical studies, Diff2SP outperforms vanilla diffusion and GAN variants in both distributional accuracy and decision quality, consistently achieving lower regret and better preservation of statistical structure. Real data experiments on two years of renewable energy scheduling confirm that Diff2SP preserves temporal dependence and cross variable correlations among wind, solar and load data, reducing expected operating cost by up to fifteen percent relative to competing methods. These results position Diff2SP as a practical and principled tool for scenario generation in stochastic optimization.

5 - Layout Optimization for a Large-Scale Grid-Connected Solar Power Plant

Chong Wang, Huazhong University of Science and Technology, Wuhan, China, People's Republic of

A solar power plant provides green electricity to the public via a power grid. As governments worldwide aim to reduce carbon emissions and achieve carbon neutrality, large-scale grid-connected solar power plants are booming. Developing such a plant requires significant investment, much of which covers construction costs. These costs, together with energy yield, critically depend on the plant's layout. Layout planning involves complex optimization problems such as district partitioning, PV component location, and cable routing. These problems are highly challenging due to large-scale instances, complex design principles, and physical constraints, and have received limited attention in the literature. Motivated by collaborative projects with an electrical engineering company in China, this paper focuses on the integrated location and routing (ILR) problem, involving locating service ways, inverters, combiner boxes, and routing cables to connect them. We develop exact algorithms to solve the ILR problem via a decomposition framework, yielding a variant of Benders decomposition (BD) proven to produce an optimal solution. An exact branch-and-cut scheme is designed to solve each subproblem by incorporating cutting planes and separation algorithms. Our approach is evaluated on 50 real-world data instances through extensive numerical experiments. Compared with the manual greedy heuristics used in practice, our method reduces total cost by about 20%. The decomposition method achieves an average gap of 0.02% between obtained lower and upper bounds, much smaller than the 16.08% gap achieved with traditional BD.

TC33

Building B Level 2 B207

Modeling for Mitigation: Designing Disease Control Protocols

Invited Session

Health Applications Society

Chair: Su Li, University of Toronto, Toronto, ON, M5G 0C5, Canada

Co-Chair: Marwan Shams Eddin, George Mason University, Euless, 76040, United States

1 - Real-Time Pandemic Control: A Digital Twin Framework for Adaptive Interventions

Osman Ozaltin, North Carolina State University, Raleigh, NC, United States, Sebastian Rodriguez Cartes, Maria Mayorga, Kevin Flores, Julie Swann

This study investigates the dynamic application of interventions to control infectious disease outbreaks, with a specific focus on mask mandates. We introduce a digital twin approach, wherein a complex agent-based model serves as the real system, and its digital twin is a stochastic compartmental model of infectious disease spread. A Bayesian methodology is utilized to update the parameters of the compartmental model as new disease spread information becomes available from the agent-based model. Subsequently, the digital twin is employed for mask mandate policy planning, and the resulting policies are implemented in the agent-based model within a continuous feedback loop. We present computational results evaluating the performance of different planning algorithms, including a grid-based threshold search approach and reinforcement learning.

2 - Optimal Lockdown Timing During An Epidemic

Noura El Hassan, American University of Beirut - Mediterraneo, Pafos, Lebanon, Bael Maddah, Fouad Ben Abdelaziz

Our research uses a Markov Decision Process to determine the optimal time to impose a lockdown, as a careful trade-off between health-related and economic aspects should be considered. We assume that the pandemic evolves based on a well-known Susceptible-Infected-Susceptible epidemic model. We show that the value function is

concave in the number of infections, which implies an optimal policy based on threshold values on the number of infections. Accordingly, we propose an efficient search and storage method that allows handling real-world problems with a large population size reaching one million within a manageable CPU time. Numerical experiments show that our policy outperforms benchmark policies. Moreover, in a case study of the Island of Corsica, we find that our policy produces favorable health recommendations.

3 - An Optimization-Based Framework to Dynamically Schedule Hospital Beds in A Pandemic

Marwan Shams Eddin, George Mason University, Euless, TX, United States

Emerging pandemics and peak infection periods can overwhelm healthcare systems, especially under limited resource availability such as hospital beds, ventilators, and ICU capacity. This strain may lead to elevated mortality rates and increased healthcare costs. In this paper, we present a novel optimization-based framework that dynamically schedules hospital beds to minimize total healthcare costs—primarily composed of patient rejection costs and associated logistical expenses—while operating under strict resource constraints. Our model accounts for multiple sources of bed supply, including standard hospital beds, insitu hospital beds, and buffer beds, and incorporates flexible resource sharing to better accommodate patient demand. Recognizing the computational complexity of the resulting optimization model, we propose a reformulation that significantly reduces run-time. We also analyze key structural properties and derive optimal solutions under specific conditions. To account for demand uncertainty, we extend the framework by integrating an SEIRD model to simulate patient demand in future pandemics and to inform proactive bed scheduling. To evaluate the practical utility of our approach, we conduct a case study based on the COVID-19 pandemic in the Northern Virginia (NOVA) region. Our results demonstrate that the proposed framework can reduce total healthcare costs by more than 55%, highlighting its potential in guiding resource allocation and preparedness strategies during health crises.

4 - Decoding State-Relevance Weights in Dual Alp: a Patient Scheduling Case Study

Su Li, University of Toronto, Toronto, ON, Canada

Markov Decision Processes serve as a foundational framework for decision-making under uncertainty. However, in high-dimensional problem settings, obtaining exact solutions is impractical due to the exponential size of the state space (the curse of dimensionality). Approximate Linear Programming (ALP) is a commonly used technique for approximating the value function with a conical combination of basis functions. A critical aspect of the ALP approach is the selection of state-relevance weights, which can markedly influence the performance of the resulting policy. In this work, we present a novel perspective on state-relevance weights through Dual Approximate Linear Programming. We first demonstrate how the choice of state-relevance weights can affect policy performance. We then propose an approach for simultaneously computing state-relevance weights and approximate value function coefficients within the ALP framework. We prove several structural properties and numerically assess the quality of our approach using applications from queueing and patient scheduling.

5 - Assessing Thresholds of Resistance Prevalence at Which Empiric Treatment of Gonorrhea Should Change Among Men Who Have Sex with Men in the US: A Cost-Effectiveness Analysis

Xuecheng Yin, Oklahoma State University, Tulsa, OK, United States

Gonorrhea is typically treated without knowing whether the infection is resistant to the antibiotic used, since common tests do not check for drug resistance. As resistance to an antibiotic increases, treatment guidelines usually switch to a different drug once a certain resistance level (like 5%) is reached. Using a lower threshold (e.g., 1%) may improve treatment success but could lead to more frequent changes in recommended antibiotics. The effects of different thresholds on gonorrhea incidence, treatment failures, costs, and quality of life are not well understood.

We built a mathematical model to study the effects of different antibiotic switch thresholds (1% to 10%) on health and economic outcomes in men who have sex with men (MSM) in the U.S. The model considered infection spread, resistance, treatment, symptoms, long-term health issues, costs, and quality-adjusted life-years (QALYs). We summarized outcomes using net health benefit (NHB), which combines cost and QALY loss. When only three antibiotics are available over 50 years (two for initial treatment, one for retreatment), changing the switch threshold had little impact on overall gonorrhea cases, costs, or QALY losses. However, if a new antibiotic becomes available, a lower switch threshold (1–2%) may improve NHB especially if drug susceptibility testing (DST) is used to guide retreatment after a failed first therapy.

Switching first-line antibiotics at lower resistance levels may not always change outcomes, but it could be beneficial if new antibiotics and DST are available. Our model focused on MSM and did not include vaccines or DST for initial treatment.

TC34

Building B Level 2 B208

Advanced Optimization methods in Data Science

Invited Session

OPT: Machine Learning

Chair: Salar Fattahi, University of Michigan, Ann Arbor, MI, United States

1 - Fast and Accurate Estimation of Low-Rank Matrices from Noisy Measurements via Preconditioned Non-Convex Gradient Descent

Hong-Ming Chiu, University of Illinois Urbana-Champaign, Urbana, IL, United States, Gavin Zhang, Richard Zhang

Non-convex gradient descent is a common approach for estimating a low-rank matrix from noisy measurements, because it has per-iteration costs as low as linear time, and is in theory capable of converging to a minimax optimal estimate. However, the practitioner is often constrained to just tens to hundreds of iterations, and the slow and/or inconsistent convergence of non-convex gradient descent can prevent a high-quality estimate from being obtained. Recently, the technique of preconditioning was shown to be highly effective at accelerating the local convergence of non-convex gradient descent when the measurements are noiseless. In this talk, we describe how preconditioning should be done for noisy measurements to accelerate local convergence to minimax optimality. For the symmetric matrix sensing problem, our proposed preconditioned method is guaranteed to locally converge to minimax error at a linear rate that is immune to ill-conditioning and/or over-parameterization.

2 - On Strong Formulations for 0-1 Fractional Programming via Disjunctive Arguments

Soo bin Choi, University of Southern California, Los Angeles, CA, United States, Andres Gomez

In this study, we propose stronger convex relaxations for 0-1 fractional programs by leveraging disjunctive programming techniques. Specifically, we derive tight upper bounds for linear fractional functions over binary domains by combining disjunctive arguments with polymatroid inequalities. Additionally, we introduce a new family of second-order cone inequalities that provide stronger lower bounds than those established in the literature. Numerical experiments indicate that the proposed formulations yield tighter relaxation gaps than existing approaches.

3 - Learning Dynamics of Deep Matrix Factorization Beyond the Edge of Stability

Soo Min Kwon, University of Michigan, Ann Arbor, MI, United States

Deep neural networks trained using gradient descent with a fixed learning rate often operate in the regime of "edge of stability" (EOS), where the largest eigenvalue of the Hessian equilibrates about the stability threshold. In this work, we present a fine-grained analysis of the learning dynamics of (deep) linear networks (DLNs) within the deep matrix factorization loss beyond EOS. For DLNs, loss oscillations beyond EOS follow a period-doubling route to chaos. We theoretically analyze the regime of the 2-period orbit and show that the loss oscillations occur within a small subspace, with the dimension of the subspace precisely characterized by the learning rate. The crux of our analysis lies in showing that the symmetry-induced conservation law for gradient flow, defined as the balancing gap among the singular values across layers, breaks at EOS and decays monotonically to zero. Overall, our results contribute to explaining two key phenomena in deep networks: (i) shallow models and simple tasks do not

always exhibit EOS; and (ii) oscillations occur within top features. We present experiments to support our theory, along with examples demonstrating how these phenomena occur in nonlinear networks and how they differ from those which have benign landscape such as in DLNs.

4 - Fastest Algorithm, in Theory and Practice, for Densest Subgraph and Other Ratio Problems: the Incremental Parametric Cut

Dorit Hochbaum, UC Berkeley, Berkeley, CA, United States

The maximum density subgraph problem is one of ratio problems that are often employed in community detection, clustering, or identifying patterns in large scale data sets, or networks. Research on the problem, in the last couple of decades, has been focused on developing heuristic methods for solving the maximum density problem claiming that flow computations are computationally prohibitive for large scale problems. These heuristics are designed for speed while sacrificing optimality. We present here a new algorithm called, the incremental parametric cut algorithm (IPC), that solves optimally the maximum density problem, as well as other "monotone integer" ratio problems related to the conductance, Cheeger constant, graph expander and normalized cut problems. The theoretical complexity of IPC is that of a single minimum cut, matching the complexity of the previously known parametric cut method. Yet in practice, IPC is dramatically faster. This is due to targeting one specific "breakpoint" solution, whereas the parametric cut procedure generates all "breakpoints". Not only does IPC overcome the computational limitations of previous cut-based approaches, but it also outperforms the state-of-the-art heuristic approaches, both in terms of solution quality, and in terms of speed, by several orders of magnitude.

TC35

Building B Level 2 B209

Managing Infectious Diseases

Invited Session

Public Sector OR

Chair: Qi (George) Chen, London Business School, London, United Kingdom

1 - The Sooner, The Better? Optimal Vaccination Policy with Limited Vaccine Supply

Qi (George) Chen, London Business School, London, United Kingdom, Miao Bai, Cuihong Li

We study the optimal single-dose vaccination policy during an infectious disease outbreak, considering both limited vaccine supply and imperfect efficacy, which provides partial immunity to each vaccinated individual. The inclusion of imperfect efficacy introduces an additional compartment to the celebrated Susceptible-Infectious-Recovered (SIR) model, giving rise to an infinite horizon nonlinear optimal control problem. We derive a closed-form expression for the optimal vaccination policy under infinite administrative capacity and establish theoretical structures of the optimal policy under finite capacity. Our results suggest that delaying the start of the vaccination process may be optimal, especially when the vaccine is less effective, the vaccine supply is more limited, and the disease is more infectious. We identify key terms as an integral part of the analysis that characterize the individual-level benefit of vaccination and, importantly, provide insights into the intuition behind vaccination delay. Building on these theoretical findings, our numerical study verifies these insights based on sensitivity results. We demonstrate the significant benefit of delay in reducing the total number of infections compared to policies without delay. Furthermore, we extend our analysis to alternative model dynamics and objective functions, confirming that our key findings remain robust and our methodology applicable across different settings. Our study contributes to the methodological framework to analyze optimal vaccination policies for infectious disease control, and uncovers an important policy structure -- vaccination delay -- that has been overlooked in the literature but has important implications for practice.

2 - Impact of Adding Cold Chain Equipment on Vaccination Timeliness

Rohan Bapat, Indian School of Business, Hyderabad, India, Sarang Deo, Sripad Devalkar

Delays in childhood vaccination remain a persistent health challenge in many low-resource settings, prolonging susceptibility to vaccine preventable diseases, and thereby undermining the effectiveness of vaccination programs. These delays stem from a combination of demand-side factors and supply-side constraints, where the former could be caregiver unavailability, child illness, or travel, while the latter includes the limited capacity of cold chain infrastructure, especially at the last mile. Inadequate last-mile cold chain capacity often necessitates a periodic or intermittent vaccination model, where vaccines are only administered on pre-determined session days, further delaying the time to vaccination. Recent multilateral initiatives to improve vaccination outcomes have focused on strengthening cold chain infrastructure through increased investments in cold chain equipment. Depending on how these investments are deployed, they can either *expand* inventory capacity at existing cold chain points or *extend* cold chain infrastructure closer to communities, potentially supporting a continuous vaccination delivery model. This research investigates how changes to cold chain infrastructure affect delays in vaccination by developing an analytical model that captures demand-side and supply-side contributors to vaccination delay. By modeling impact of these two distinct investment strategies - *expansion* and *extension* of cold chain capacity, we assess how such infrastructure improvements influence the likelihood of children receiving vaccines at age-appropriate times. Our findings offer operational insights to guide the design and allocation of vaccination program investments to improve timely vaccination.

3 - Mobile Vaccination Team Weekly Scheduling and Vehicle Routing

Jinghong Miao, University of Southern California, Los Angeles, CA, United States, Jing Jin, Sze-chuan Suen, Maged Dessouky, Fernando Ordonez

Every week, planners at the Los Angeles Department of Public Health (LADPH) Mobile Vaccination Team must route and schedule nurses to geographically dispersed vaccination sites for the following week. Each vaccination event has staff requirements, in addition weekly nurse working-hour limits and event time windows must be observed. This problem combines both aspects of scheduling and routing problems, resulting in a formulation of intersecting vehicle routing problems (VRPs) that must be solved simultaneously. We formulate a mixed-integer linear program that simultaneously determines nurse assignments and travel routes over a one-week horizon. The objective minimizes total travel distance while considering a fairness metric that caps workload disparities. However, standard solvers cannot identify feasible solutions within operational computational times for realistic problems at scale, so we propose a heuristic framework that 1) discretizes time into intervals to prune infeasible solutions, 2) partitions the full horizon into subproblems, and 3) iteratively refines solutions and bounds. Computational experiments on both simulated and real instances demonstrate that our heuristic provides a better solution than standard solvers in significantly less computation time. This work provides a scalable method for nurse scheduling and routing under fairness considerations.

4 - Optimized Hiv/Aids Resource Allocation in Ohio: a Linear Programming Approach

Godfred Ahenkroa Kesse, University of Central Florida, Orlando, FL, United States

This study employs a linear and integer programming approach to optimize HIV resource allocation in Ohio, aiming to minimize new infections and enhance the impact of limited resources. With the advances in HIV prevention and treatment, Ohio faces challenges in addressing disparities in access to healthcare, particularly among high-risk populations. The proposed model integrates data on infection rates, transmission patterns, demographic factors, and cost-effectiveness to provide a decision-support framework for policymakers. Using epidemiological data and equity constraints, the model prioritizes high-risk regions and populations while ensuring fair resource distribution. Results indicate that increased funding allocations significantly enhance the potential to avert new infections. This research underscores the importance of data-driven strategies in achieving the objectives outlined in the HIV National Strategic Plan, emphasizing targeted interventions and equitable healthcare access to curb the epidemic effectively.

5 - Antigenicity Prediction and Vaccine Recommendation of Human Influenza Virus A (H3n2) Using Convolutional Neural Networks

Eva Lee, The Data and Analytics Innovation Institute, Atlanta, GA, United States

The rapid and large-scale pathogenesis of influenza virus requires constant monitoring and frequent vaccine development to protect the world population not only from seasonal influenza but also from novel influenza A viruses that could trigger a pandemic. Analytical and machine learning models have been applied to facilitate the process of antigenicity determination. In this study, we designed deep convolutional neural networks to predict Influenza antigenicity. We first systematically compared the all 20 amino acid properties and amino acid substitution matrices for their predictability. We then optimized the structure of the convolutional neural networks using particle swarm optimization. The optimal neural networks outperform other predictive models with a validation accuracy of 0.958. Further, the results show that our model

outperforms other existing models, and could potential improve the vaccine recommendation process. Overall, we introduce a comprehensive pipeline for optimizing a convolutional neural network in the modeling of Influenza A's antigenicity and vaccine recommendation.

TC36

Building B Level 2 B210

Adaptive Stochastic Nonlinear Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Qi Wang, Lehigh University, Bethlehem, PA, United States

1 - Adaptive Fixed-Step Stochastic Gradient

Raghu Pasupathy, Purdue University, West Lafayette, IN, United States, Akshita Gupta, Raghu Bollapragada, Nung Kwan Yip

A folklore adaptive fixed-step stochastic gradient method works as follows. Execute stochastic gradient with a fixed step until the iterates approach stationarity; now re-execute stochastic gradient starting from the final iterate of the previous execution and with the revised fixed step until the iterates again approach stationarity, and so on. This restart scheme is attractive on two accounts: (i) it is simple to implement, and (ii) due to the Markov structure, each execution of fixed-step stochastic gradient generates iterates that (provably) attain stationarity exponentially fast. The first version of this scheme seems to have been conceived by G. Pflug in 1983, and several prominent authors including U. Shanbhag, J. Nocedal, and L. Xiao have worked or commented on enhancements, with some reporting strikingly positive numerical results. Considering this empirical success, we present a unified view of adaptive fixed-step gradient methods using *statistical testing functionals*, reflecting our viewpoint that most if not all adaptive fixed-step methods are explicitly or implicitly attempting to detect stationarity of iterates as a signal for reducing step size. The unification we propose facilitates characterizing generic consistency and optimal complexity conditions for adaptive fixed-step gradient methods, while also providing a template for more general settings. Time permitting, we will present a specific variation that is robust to tuning and that has performed favorably against well-tuned fixed-step SGD algorithms on standard problems.

2 - On the Resolution of Stochastic Mpecs over Networks: Distributed Implicit Zeroth-Order Gradient Tracking Methods

Mohammadjavad Ebrahimi, Rutgers University, Piscataway, NJ, United States, Uday Shanbhag, Farzad Yousefian

The mathematical program with equilibrium constraints (MPEC) is a powerful yet challenging class of constrained optimization problems, where the constraints are characterized by a parametrized variational inequality (VI) problem. MPECs subsume broad subclasses of bilevel optimization and Stackelberg games, which have numerous applications in power markets, transportation networks, and structural design problems. While efficient algorithms for addressing MPECs and their stochastic variants (SMPECs) have been recently presented, distributed SMPECs over networks remain unexplored. This work aims to develop fully iterative methods with complexity guarantees for resolving distributed SMPECs in two problem settings: (1) distributed single-stage SMPECs and (2) distributed two-stage SMPECs. In both cases, the global objective function is distributed among a network of agents that communicate cooperatively, and the implicit objective is possibly nondifferentiable and nonconvex. We propose single-stage and two-stage zeroth-order distributed gradient tracking optimization methods where the gradient of a smoothed implicit objective function is approximated using two (possibly inexact) evaluations of the lower-level VI solutions. For the exact settings, we achieve the best-known complexity bound for centralized nonsmooth nonconvex stochastic optimization, that is $\mathcal{O}(\epsilon^{-\frac{3}{2}})$ where n denotes the upper-level decision dimension. This complexity bound is also achieved for the inexact setting of the distributed two-stage SMPEC. In addressing the inexact setting of the single-stage problem, we derive the overall complexity of $\mathcal{O}(\epsilon^{-\frac{7}{2}})$, improving the dependence on n compared to the existing results for the centralized SMPECs. Preliminary numerical experiments demonstrate the efficiency of our methods across various network topologies compared to centralized zeroth-order schemes.

3 - High Probability Complexity Bounds of Trust-Region Stochastic Sequential Quadratic Programming with Heavy-Tailed Noise

Yuchen Fang, University of California, Berkeley, Berkeley, CA, United States, Javad Lavaei, Sen Na

In this paper, we consider nonlinear optimization problems with a stochastic objective and deterministic equality constraints. We propose a Trust-Region Stochastic Sequential Quadratic Programming (TR-SSQP) method and establish its high-probability iteration complexity bounds for identifying first- and second-order ϵ -stationary points. In our algorithm, we assume that exact objective values, gradients, and Hessians are not directly accessible but can be estimated via zeroth-, first-, and second-order probabilistic oracles. Compared to existing complexity studies of SSQP methods that rely on a zeroth-order oracle with sub-exponential tail noise (i.e., light-tailed) and focus mostly on first-order stationarity, our analysis accommodates irreducible and heavy-tailed noise in the zeroth-order oracle and significantly extends the analysis to second-order stationarity. We show that under heavy-tailed noise conditions, our SSQP method achieves the same high-probability first-order iteration complexity bounds as in the light-tailed noise setting, while further exhibiting promising second-order iteration complexity bounds. Specifically, the method identifies a first-order ϵ -stationary point in $O(\epsilon^{-2})$ iterations and a second-order ϵ -stationary point in $O(\epsilon^{-3})$ iterations with high probability, provided that ϵ is lower bounded by a constant determined by the irreducible noise level in estimation. We validate our theoretical findings and evaluate the practical performance of our method on CUTEst benchmark test set.

4 - A Stochastic Linesearch Method (Slam) for Minimizing Expectation Residuals

Qi Wang, University of Michigan, Ann Arbor, Ann Arbor, MI, United States, Uday Shanbhag, Yue Xie

Most existing rate and complexity guarantees for stochastic gradient methods in L -smooth settings impose restrictive requirements on steplength sequences, mandating that such sequences be non-adaptive, non-increasing, and abide by an upper bound given by a/L for some positive scalar a . This severely limits the applicability and adversely affects the performance of such schemes for three key reasons: (i) requires knowing L ; (ii) precludes larger steps when L is large; and (iii) employs non-increasing steplength sequences. Motivated by these shortcomings, we present an Armijo-enabled stochastic linesearch framework with standard stochastic zeroth- and first-order oracles. The resulting steplength sequences are not non-increasing, while enjoying a mild bound related to L . We refer to the proposed scheme as quasi parameter-free since by appropriate selection of algorithm parameters, the dependence on L is seen to be almost insignificant. We then prove that the expected stationarity residual diminishes at a rate of $\mathcal{O}(1/\sqrt{K})$, where K denotes the iteration index. Furthermore, the resulting iteration and sample complexities for computing an ϵ -stationary point are $\mathcal{O}(\epsilon^{-2})$ and $\mathcal{O}(\epsilon^{-(4+\delta)})$, where $\delta > 0$ is a fixed small number. Analogous guarantees are provided in the PL setting as well as in convex regimes. Preliminary numerical experiments are promising as the performance of the scheme competes well with de-facto choices such as Adam.

5 - An Adaptive Augmented Lagrangian Method for Deterministic and Stochastic Non-Convex Optimization

Tianzhu Liu, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Michael O'Neill

We present an inexact Augmented Lagrangian algorithm for solving non-linear, non-convex optimization problems. Unlike most recently proposed Augmented Lagrangian methods with worst-case complexity guarantees, we utilize adaptive penalty parameter updates and full dual step sizes. We show that the method matches the best-known worst-case complexity results for Augmented Lagrangian methods (up to logarithmic factors) when both the function and constraints are deterministic, when the function is stochastic and the constraints are deterministic, and when both are stochastic. Experiments on CUTEst test problems confirm the practical advantages of the proposed approach over Augmented Lagrangian methods with non-adaptive penalty parameters and/or short dual step sizes in the deterministic setting. Numerical results on stochastic constrained optimization problems in machine learning also confirm these findings.

6 - Riemannian Proximal Sampler for High-Accuracy Sampling on Manifolds

Yunrui Guan, Rice University, Houston, TX, United States, Krishnakumar Balasubramanian, Shiqian Ma

We introduce the Riemannian Proximal Sampler, a method for sampling from densities defined on Riemannian manifolds. The performance of this sampler critically depends on two key oracles: the Manifold Brownian Increments (MBI) oracle and the Riemannian Heat-kernel (RHK) oracle. We establish high-accuracy sampling guarantees for the Riemannian Proximal Sampler, showing that generating samples with ϵ -accuracy requires $O(\log(1/\epsilon))$ iterations in Kullback-Leibler divergence assuming access to exact oracles and $O(\log(1/\epsilon))$ iterations in the total variation metric assuming access to sufficiently accurate inexact oracles. Furthermore, we present

practical implementations of these oracles by leveraging heat-kernel truncation and Varadhan's asymptotics. In the latter case, we interpret the Riemannian Proximal Sampler as a discretization of the entropy-regularized Riemannian Proximal Point Method on the associated Wasserstein space. We provide preliminary numerical results that illustrate the effectiveness of the proposed methodology.

TC37

Building B Level 2 B211

Scenario-Based Decomposition for Stochastic Programming: Methods and Software

Invited Session

OPT: Optimization Under Uncertainty

Chair: Jean-Paul Watson, Lawrence Livermore National Laboratory, Livermore, CA, 94550, United States

1 - Large-Scale Experimentation and Analysis of Hpc-Based Scenario Decomposition via Progressive Hedging with Mpi-Sppy

Jean-Paul Watson, Lawrence Livermore National Laboratory, Livermore, CA, United States, David Woodruff, Bernard Knueven

Parallel implementations of scenario-based decomposition strategies are now scalable to thousands and millions of scenarios, thanks to advances in modeling systems (e.g., Pyomo) and supporting meta-solvers (e.g., mpi-sppy). We describe challenges and their solution considering a large-scale power grid capacity expansion model, where scenarios represent individual days of weather. We discuss in detail various algorithmic tuning and approaches to rapidly finding high-quality solutions to these stochastic programs, utilizing the mpi-sppy meta-solver library.

2 - What's New in the MIP-SPPY Solver Library

David Woodruff, UC Davis, Davis, CA, United States, Bernard Knueven, Jean-Paul Watson

MIP-SPPY is software for stochastic programming that can run on a laptop, but uses MPI (message passing interface) to perform well on distributed memory computers. In this talk I will talk about high level changes to the user interface and low level changes to the way we manage interaction between algorithms running asynchronously in parallel.

3 - Primal and Dual Progressive Hedging Heuristics for Stochastic Mixed-Integer Programs

Bernard Knueven, National Renewable Energy Laboratory, Golden, CO, United States, Jean-Paul Watson, David Woodruff

The *mpi-sppy* meta-solver library facilitates scalable solutions to large-scale stochastic programming problems through the asynchronous execution of primary algorithms, such as Progressive Hedging, and assisting algorithms that compute objective value bounds or accelerate convergence of the primary algorithm. In this talk, we present recent enhancements to *mpi-sppy*, including newly developed assisting algorithms aimed at improving both primal and dual convergence of Progressive Hedging. We evaluate the effectiveness of these heuristic strategies on a challenging, large-scale power grid capacity expansion model.

4 - Progress Report: Toward Asynchronous Parallel Computation of Lagrangian Bounds for Multi-Stage Stochastic Mixed-Integer Programs

Jonathan Eckstein, Rutgers University, Piscataway, NJ, United States, Jean-Paul Watson, David Woodruff

We describe an effort to compute Lagrangian bounds for multi-stage stochastic MIPs in which scenario/bundle subproblems are solved in parallel through essentially continuous computational flows without synchronization gaps. The technical approach is to use the Frank-Wolfe-like simplicial decomposition (SD) approach of Boland et al., but in the context of a projective splitting procedure rather than standard synchronous progressive hedging. This combination is not entirely straightforward and involves some adjustments to the SD procedure (we may also discuss some enhancements to SD for this application). The implementation combines two levels of parallelism: a C++ native multithreading lower level and an upper level employing nonblocking MPI reductions.

TC38

Building B Level 2 B212

Recent Progress in Optimization Software III

Invited Session

OPT: Computational Optimization and Software

Chair: Hans Mittelmann, Arizona State University, Tempe, AZ, United States

1 - The Latest Developments in The Knitro Optimization Solver

Richard Waltz, Artelys, Los Angeles, CA, United States

Knitro was originally developed in the 1990s as an interior-point algorithm for general nonlinear, non-convex optimization. Over the years, Knitro evolved into a more general optimization toolbox that includes an Active-Set LP-based solver, a Sequential Quadratic Programming (SQP) solver, specialized LP, QP, and SOCP solvers, a branch-and-bound solver for mixed integer programming, and multi-start heuristics for global optimization. To add to this toolbox of algorithms, we have recently started developing first-order methods that do not require any matrix factorizations. The hope is that these might be able to provide useful solutions to extremely large-scale models where the factorizations in interior-point methods become too expensive. In this talk we will present some of this work, primarily based on Augmented Lagrangian (AL) type methods, as well as present other updates from the latest Knitro 15.0 release.

2 - The Sas Optimization Solvers: Recent Improvements and Future Developments

Rob Pratt, SAS, Cary, NC, United States, Philipp Christophel, Menal Guzelsoy, Laszlo Ladanyi, Bochuan Lyu, Amar Narisetty

In this talk we present the recent improvements to the SAS linear and mixed integer linear optimization solvers and discuss future developments.

3 - Recent Progress in the Global Optimization of NLPs and MINLPs with Baron

Nikolaos Sahinidis, Georgia Tech, Atlanta, GA, United States, Yi Zhang

We review historical progress and recent developments toward the global optimization of nonlinear and mixed-integer nonlinear programming problems with BARON. We discuss relaxation construction, preprocessing, branching, and tree management techniques. We present results for a large test set and emphasize recent progress, especially for quadratic and integer programs.

4 - Recent Advances in the OSQP Solver

Ian McInerney, Imperial College London, London, United Kingdom, Amit Solomon, Vineet Bansal, Bartolomeo Stellato

OSQP is a fast, light, and widely popular open-source solver for quadratic optimization based on the alternating direction method of multipliers. OSQP includes a code-generation facility to generate a light embeddable version of the solver, and the ability to compute derivatives of its solution with respect to the problem parameters. It uses a modular linear algebra framework to seamlessly run on three different platforms: default single-threaded for embedded applications, the Intel Math Kernel Library for multicore CPUs, and CUDA for GPUs. In this talk, we present the latest changes, improvements, and additions to OSQP. These include algorithmic improvements such as duality-gap termination criteria and step-size selection, improved GPU performance, and better integration with the language ecosystems of the three major high-level wrappers: Python, Julia and MATLAB.

5 - Smoothie: A Highly Parallel Ensemble Solver

Yuji Shinano, Zuse Institute Berlin, Berlin, Germany

The Ubiquity Generator (UG) Framework is a software framework to parallelize state-of-the-art solvers. Based on UG, SCIP and its customized solvers for Steiner tree problems (SCIP-Jack), QUBOs (Qubowl), and Pseudo-Boolean optimization, are parallelized. The UG framework has been generalized in the upcoming version 1.0.0 so that the parallelization controlling mechanism can be highly customized. The feature allows us to develop ensemble solvers, in which several different solvers run in parallel and communicate information. An experimental implementation of Smoothie (Solver Mixing Object Oriented Hybrid Integrated Executable), in which SCIP and HiGHS can run in parallel, has been developed. In this talk, we present the latest status of Smoothie where commercial solvers Xpress and Gurobi are used jointly to solve hard MIP instances.

6 - Latest Results for the Xopt Solver

Xiaodong Luo, Chinese University of Hong Kong, ShenZhen Campus, ShenZhen, Guangdong province, China, China, People's Republic of

In this talk, we are introducing XOPS LP/MILP solver. We have added Objective splitting, max cut primal heuristic and distributed solving framework since last year. We will report on our automatic model building environment as well.

TC39

Building B Level 2 B213

Leveraging Structure in Integer Programs: Decomposition, Symmetry, and Parameterization

Invited Session

OPT: Integer and Discrete Optimization

Chair: Santanu Dey, ISyE Georgia Tech, Atlanta, GA, United States

Co-Chair: Aleksandr Kazachkov, University of Florida, Gainesville, United States

1 - Symmetries and Lift-and-Project Operators

Yuri Faenza, Columbia University, New York, NY, United States, Victor Verdugo, José Verschae, Matias Villagra

Symmetries are functions that preserve an object's structure while mapping it onto itself. They are prevalent in mathematics, and integer programming is no exception -- at least 47% of mixed-integer linear programs in MIPLIB 2010 present some symmetry. It is now folklore that symmetric integer programs are difficult to solve: branch-and-cut algorithms typically show limited improvement in the dual bound when applied to such problems.

With the goal of further understanding why symmetries can slow down solvers, we theoretically investigate how various lift-and-project operators (Sherali-Adams, Lovasz-Schrijver, etc.) perform on polytopes as a function of properties of the group of symmetries acting on them (i.e., no specific structure of the polytopes is assumed). Our results show a hierarchy collapse: lift-and-project operators which are known in general to perform differently behave, will under certain conditions, be as (in)effective for improving LP bounds of symmetric linear programs.

2 - Random 2-Stage Stochastic Packing-Ips Can be Solved in Poly-TIME Using Decomposition Algorithm

jingye xu, Georgia Institute of Technology, Atlanta, GA, United States

Decomposition algorithms are widely used to solve linear integer programs (LIPs) with block structures. Classical methods like the Lagrangian decomposition or the Benders' method typically settle original problems by solving a sequence of smaller-dimensional LIPs over each block with possibly additional cutting planes. Although those decomposition algorithms are able to solve large-scale LIPs with block structures efficiently, the theoretical understanding of those methods is relatively weak. The existing convergence guarantees are all exponential in the number of variables in each block, which contradict their good practical performance. In this paper, our goal is to understand the behavior of a variant of Lagrangian decomposition recently introduced by Cifuentes, Dey, Xu (IPCO, 2025). When applied to random two-stage LIPs where the entries of the coefficient matrix and the objective function are randomly sampled, we show that such decomposition method solves the problem in poly-iteration when the number of constraints of each block is fixed.

3 - Parametric Polyhedra in Mixed-Integer Programming

Diego Moran, Rensselaer Polytechnic Institute, Troy, NY, United States, Gustavo Angulo, Silvia Di Gregorio

We present some old and new results on arbitrary families of parametric polyhedra. First, if the constraint matrix is fixed, in the literature there are structural results for the integer hull and the finiteness of cutting plane closures for varying r.h.s. For instance, recently, Becu et al. proved in "Approximating the Gomory Mixed-Integer Cut Closure Using Historical Data" that the GMI closure of this family is finitely generated, in the sense that there exists a finite list of aggregation weights defining the GMI cuts that give the GMI closure for any polyhedra in the family. We extend this result for other cutting plane closures. Second, if the family of parametric polyhedra is arbitrary but all polyhedra in the family have the same integer hull, they define the same MIP, and we can leverage this information to understand and solve MIPs better. These families have been used to understand theoretical properties of the rank of cutting planes and to obtain better formulations. We present an application of these same-integer-hull families to formulations for the Asymmetric Traveling Salesman Problem.

4 - Tensor Completion via Integer Programming

Chen Chen, The Ohio State University, Columbus, OH, United States, Alexander While, Anil Aswani, William Caballero, Nathan Gaw

Tensor completion is a general-purpose model for imputing missing or unobserved entries, typically under the assumption of underlying low-rank structure. We present new results on our approach, which leverages a linear number of calls to an integer programming oracle to ensure convergence to a low-rank solution. Although modestly slower than common approaches, our approach attains the information-theoretic bound on minimum error with respect to sample size. Latest results include GPU acceleration, as well as an application to power systems.

5 - Scaling the Simple Disjunctive Cut Principle

Aleksandr Kazachkov, University of Florida, Gainesville, FL, United States, Chen Chen, Connor Johnston, Diego Moran

We introduce a new perspective on computing facet-defining inequalities for the convex hull of a union of polyhedra, where each term in this disjunction is defined by a single linear inequality imposed on a (shared) simple cone. The classic approach to generating these inequalities is a linear program from a reverse polar formulation. We avoid such optimization and instead provide a tree-based enumeration algorithm through a geometric interpretation of scaling the disjunctive cut principle, which computes inequalities with a simple formula. Our results tighten the exponential upper bound from the literature on the number of extra constraints needed to obtain the convex hull, and we show complementary families of instances in which quadratically-many inequalities are needed.

TC40

Building B Level 2 B214

Service Science I

Contributed Session

Contributed

Chair: Tana Wang, Shanghai Jiao Tong University, N/A

1 - Innovation Efficiency Across Product, Service, and Combined Innovation

Jaeho Shin, Hansung University, Seoul, Korea, Republic of, Changhee Kim, Hongsuk Yang

Firms increasingly provide a combined offering of products and services, a trend often referred to as servitization or product-service systems (PSS). This approach enables firms to pursue innovation in both products and services. Based on innovation activity, this study classifies firms into three categories: (1) firms engaging in both product and service innovation, (2) firms engaging only in product innovation, and (3) firms engaging only in service innovation. We compare the innovation performance of these

three groups in terms of innovation efficiency, which accounts for both innovation outcomes and inputs such as resources and efforts. To measure innovation efficiency, we employ data envelopment analysis (DEA), a widely used non-parametric method for assessing efficiency. The empirical analysis uses firm-level data from the Korean Innovation Survey (KIS), which provides comprehensive information on innovation activities in Korea. The results may help managers identify appropriate innovation strategies by considering their available resources and innovation objectives.

2 - From Data Granularity to Model Connectivity: Emotion-Driven Customer Satisfaction Prediction in Call Center Conversations

Thomas De Cleen, Vlerick Business School, Gent, Belgium, Philippe Baecke, Frank Goedertier

Customer satisfaction is a crucial metric for evaluating call center performance, yet traditional survey methods suffer from low response rates and limited scalability leading to blind spots in service quality management. This study compares multiple frameworks that leverage utterance-level and call-level multimodal emotion recognition to predict customer satisfaction. Our pipeline processes 25008 conversations from a large utilities company, extracting textual and acoustic emotion features and feeding them into a transformer-based model with self-distillation for utterance-level emotion recognition. Customer satisfaction labels are self-reported on a 0–10 scale (mean 8.34, SD 2.59, N = 25 008).

Compared to baseline performance of weighted average emotions over the entire conversation, we test three strategies. First, predicting satisfaction from aggregated call-level emotion dynamics; second, using a many-to-one LSTM over the sequence of utterance-level emotions; and third, employing a hierarchical model that feeds utterance-level emotion embeddings directly into the LSTM. To our knowledge, this is the first study to structurally compare multiple methods of predicting satisfaction with utterance-level and call-level emotions. Early results indicate promising predictive performance.

Our framework is designed for seamless integration into existing contact-center systems: organizations that collect call transcripts, recordings, and post-call scores can train their own satisfaction predictors without any manual emotion annotation. By enabling reliable estimation for non-responders, this methodology promises to enhance service-quality monitoring and operational decision-making in real-world call centers.

3 - Context-Aware AI Recommendations: Integrating Situational and Attitudinal Dimensions

Sung Yeon Kim, Korea University, Sejong, Korea, Republic of, JAESEOK YUN, Seonghyeon Han, Jinmin Kim

Most AI recommendation systems primarily rely on users' past behaviors, often overlooking immediate situational contexts (e.g., current mood or goals) and broader attitudinal dispositions toward products or services. Such limitations can result in misaligned recommendations and decreased user satisfaction. This study proposes and empirically tests an AI recommendation model that integrates users' situational and attitudinal states, both classified along positive–negative spectrums. Through a survey study with 800 participants, we conduct a 2x2 group comparison (situation: positive/negative × attitude: positive/negative) to examine how these dimensions moderate the established relationship between information quality, system quality, recommendation quality, and user satisfaction. Prior research has shown that information and system quality influence recommendation quality, which in turn drives user satisfaction. Building on this foundation, we hypothesize that alignment with users' psychological states enhances satisfaction, with the highest satisfaction occurring under a positive situation and positive attitude, and the lowest under a negative situation and negative attitude. This study contributes a comprehensive theoretical framework that extends beyond physical contexts (e.g., location, time) to incorporate emotional and cognitive dimensions, addressing a significant gap in current research. Findings will inform the design of emotionally responsive AI recommendation systems that better align with users' dynamic psychological states and deliver more meaningful, context-sensitive user experiences. The findings of this study can also contribute to the future development of standards and conformity assessment criteria for user-centric context-aware AI recommendation systems.

4 - Adaptive Emotional Support Through Conversational System Design: The Role of Responsiveness and User Autonomy

Austin Zhiqi Wang, City University of HongKong, HongKong, Hong Kong

Digital conversational system shows potential for providing emotional support, but how its interactive design might be optimized to alleviate loneliness remains unknown. Prior research largely emphasizes static empathy mechanisms or anthropomorphic cues, while ignoring adaptive and user-driven interface elements. To address this gap, we develop and propose two novel features to renew the current conversational systems: Emotional Affordance Responsiveness (EAR) and User-Controlled Prompt Frequency (UCPF) from the system design perspective. EAR enables systems to adaptively tailor the tone and depth of prompts by detecting emotional cues in user input, for example, offering a reflective prompt after identifying sadness in sentiment. UCPF allows users to adjust prompt content during the dialogue. Integrating affective congruence theory and self-determination theory, we hypothesize that EAR enhances emotional congruence, while UCPF fosters interactional autonomy, both contributing to reduced loneliness. However, excessive user-set frequency may unintentionally result in short-term fatigue, especially among socially sensitive people. To address this issue, we propose social sensitivity as an individual difference moderator that modifies the effects of UCPF on the mediators. A randomized between-subjects experiment using an interactive conversational platform will be conducted to test the research model. Theoretically, this study enriches the design perspective on digital emotional support by identifying key socio-affective design variables. Practically, it offers insights for building digital conversational companions that are not only emotionally responsive but also customizable, thus better aligning with diverse users' preferences in loneliness interventions.

5 - Redeeming Third-Party Experiences: How Loss Aversion Influences Loyalty Point Redemption Decisions

Yiran Liu, James Madison University, Harrisonburg, VA, United States, Zheng Wu

As loyalty programs increasingly offer experiential rewards, such as concerts, sporting events, and culinary experiences, understanding consumer behavior in point-redemption settings has become more critical. While loss aversion is well studied in the context of tangible goods, its role in the experience economy remains under-explored. This study investigates how consumers' decisions are shaped when redeeming loyalty points, earned through service providers like airlines and hotels, for third-party experiential rewards, particularly when consumers exhibit loss-averse preferences. We examine two redemption mechanisms: fixed-price and auction-style. Drawing on prospect theory, we propose that each mechanism evokes two types of loss perception: financial loss (depletion of loyalty points) and anticipated regret (fear of missing a valued experience). Through controlled experiments, we explore consumer responses across two redemption mechanisms, three point commitment levels (low, medium, high) and two fear of missing out (FOMO) conditions, manipulated via perceived scarcity (low vs. high). In each scenario, participants decide whether to redeem their points immediately for a third-party experience or conserve points for future use with the original service provider. Interestingly, our findings reveal that loss-averse consumers are more likely to redeem points immediately when faced with higher point commitment levels. Moreover, auction-style redemptions further increase redemption likelihood, particularly under high FOMO conditions, as the anticipated regret of missing out outweighs concerns about financial loss. This research contributes to the behavioral operations and marketing literature on point-redemption behavior in experiential contexts and practically aid firms in loyalty program design, enhance consumer engagement, and inform policymakers in service-driven marketplaces.

6 - Matching Strategy and Product Line Design on Home Service Platforms

Hui Xiong, Donghua University, Shanghai, China, People's Republic of, Ying-Ju Chen, Lu Hsiao

In recent years, we have witnessed the growth of home service platforms. These platforms use commission-based contracts to recruit suppliers, provide multiple products to serve consumers, and match suppliers with the products. The platforms must optimize their matching strategies, commission-based contracts, and product quality decisions to improve the profit. Therefore, we build a model to investigate the matching strategies, the contract design, and the quality decisions in which suppliers with different service efficiency provide products for heterogeneous consumers. The consumers with relatively high (low) willingness to pay for quality belong to the high (low) segment. When low-segment consumers become very important (i.e., they take a large proportion, and their willingness to pay for quality is high), the platform may induce high-efficiency suppliers to offer low-quality products. Moreover, the platform is more likely to adopt this strategy when the matching probability for low-quality products is higher. However, when the heterogeneity between two segments is sufficiently small, the platform intends to induce high-efficiency suppliers to offer high-quality products and distort the quality level upward for the high segment. In this case, the quality difference between the two products increases in low-segment consumers' willingness to pay for quality.

7 - Store-Wide Shelf Space Planning in An Epidemic

Bacel Maddah, American University of Beirut, Beirut, Lebanon, Tulay Flamand, Ahmed Ghoniem

Epidemics like COVID-19 highlighted the importance of safety in public spaces. We address shelf-space allocation in a grocery store via a nonlinear integer program that allocates product categories to shelves in a way that minimizes congestion and reduce infections, while maintaining profitability from impulse buying. Queuing theory is used in estimating the congestion and health-related metrics.

8 - Home Service TIME-Dependent Routing and Appointment Scheduling: a Deep Reinforcement Learning Approach

Tana Wang, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Kaibin Wang, Guohua Wan

We investigate the Time-Dependent Routing and Appointment Scheduling (TDRAS) problem in home service systems by introducing an end-to-end decision-making framework that integrates a temporally-aware graph neural network with a deep reinforcement learning (DRL) policy to improve scheduling efficiency and adaptability. The proposed framework incorporates a learnable traffic prediction module that estimates travel durations from historical records and time-of-day patterns, supplying the model with context-sensitive temporal representations. These representations are embedded into an enhanced GINConv architecture to enable more informed and flexible node-level reasoning. A unified actor-critic network then jointly outputs discrete customer selection actions and continuous appointment time suggestions, facilitating the coordinated optimization of routing and time-slot allocation. Empirical results using real-world traffic data from Shanghai's Xuhui District show that our approach consistently outperforms traditional benchmarks—including mixed-integer programming and other heuristic algorithms—as well as solutions by large language foundation models, with respect to the weighted travelling cost, expected idle cost, and customer waiting time. This framework provides a compact, adaptive, and practically deployable solution for intelligent scheduling in time-dependently dynamic home service environments.

TC41

Building B Level 3 B301

Diverse Interventions for a Healthier World

Invited Session

Health Applications Society

Chair: Amir Karimi, UTSA, San Antonio, TX, United States

Co-Chair: Yiqi Tian, University of Pittsburgh, Pittsburgh, PA, United States

1 - Empowering Frontline Health Workers in Lmics to Manage Stock-Outs: Evidence from a Training Program in Indonesia

Amir Karimi, University of Texas at San Antonio, San Antonio, TX, United States

Through a collaboration with JSI Research & Training Institute, Inc., we gain access to fine-grained, proprietary data from approximately 17,000 health facilities across Indonesia. Leveraging this data, we evaluate the impact of a large-scale inventory management training program targeted at frontline health workers. Our research highlights the practicality and scalability of human-capital-focused interventions, such as training, in improving health outcomes in LMICs.

2 - Optimizing Surveillance Protocols for Hospital-Acquired Infections: Case of Mrsa

Esma Akgun, University of Chicago Booth School of Business, Chicago, IL, United States, F. Safa Erenay, Sibel Alumur Alev, William A. Ciccotelli

Roommates of nosocomial methicillin-resistant *Staphylococcus aureus* (MRSA) cases have a high risk of MRSA acquisition. Infection prevention and control guidelines recommend that these individuals be isolated and undergo surveillance testing; however, optimal surveillance testing and isolation strategies are unknown. We develop a stochastic dynamic programming model to optimize testing and isolation decisions for the contacts of index MRSA cases in hospitals to minimize the loss of quality-adjusted life years and the number of MRSA colonizations. Our model provides a novel framework that formulates the problem as a Markov decision process, capturing the partial observability of patients' health states by keeping sufficient information within the model state and dynamically incorporating new test results, thereby overcoming the curse of dimensionality associated with partially observable Markov decision processes. We optimally solve the model using clinical data from a Canadian community hospital and perform sensitivity analyses on key parameters, including disutility values and disease parameters. We computationally analyze the structure of the optimal testing decisions, which recommend varying the frequency and timing of tests based on initial test results and room configurations. Optimal testing decisions offer valuable insight for policymakers in improving health outcomes while reducing associated costs compared to current guidelines and practices. In addition, we compare the performance of various practical MRSA testing protocols using the proposed modeling framework and determine those with close-to-optimal performance while balancing clinical efficacy and cost-effectiveness. The proposed modeling framework can be applied to other healthcare-associated infections and infectious diseases.

3 - Optimal Assignment Guidelines for Maternity Waiting Homes in Resource-Constrained Environments

Yuhang Du, London Business School, London, United Kingdom, George Chen, Jeremie Gallien, Jody Lori

Maternity waiting homes (MWHs) are increasingly used to increase skilled birth attendance in regions with low obstetric services density. We seek to optimize their assignment policy, that is whether and when a pregnant woman should go to her local MWH given available accommodation capacity and individual characteristics including living location. Our model leverages clinical data on labor onset and latent phase duration to minimize the predicted number of unassisted deliveries resulting from such policy in the local population. We present explicit mathematical solutions for different variants of this problem and IP formulations of discretized approximations capturing practical implementation constraints. Our numerical experiments leverage detailed primary and secondary data on the population and health facilities of Bong County, Liberia. They suggest that our proposed pooling and assignment policy interventions for the existing MWH capacity in that region may plausibly halve the incidence rate of unassisted deliveries and reduce the probability of unassisted deliveries for the most vulnerable women from 70% to 15%. Implementing data-driven individual assessments of unassisted delivery risk and MWH assignment policies may thus constitute impactful and low-cost interventions in regions with limited MWH accommodation capacity.

4 - Profit Meets Wellness: How Consumer Valuation Uncertainty and Operational Cost Structure Shape Multi-Attribute Product Line Design

Iva Rashkova, Washington University, St Louis, MO, United States, Lingxiu Dong

The rise of wellness-conscious consumers has led to record demand for products with wellness attributes, such as low-sugar foods and ultraviolet-protective clothing. This market trend presents a profit-growth opportunity for established companies, which have dominated the market based on traditional attributes, such as the taste of food and the appearance of clothing. Yet, taking advantage of this opportunity is challenging due to the increased operational costs associated with delivering products with wellness attributes and the lack of information on consumers' valuation of wellness. We present a model of a monopolist developing and producing conventional and wellness products to serve a two-segment market consisting of wellness-conscious and wellness-neutral consumers. We propose a two-dimensional differentiation-contingency framework to depict the rich set of possible optimal strategies the firm could use to segment the market and explore how the market environment and the firm's operational environment affect the firm's choice of the optimal product strategy. The firm's operational cost structure further leads to different prioritization within the wellness product's quality dimensions: low development cost (resp. low coupling cost between the two quality dimensions) induces prioritization of the wellness (resp. traditional) quality of a wellness product. Examining consumer wellness consumption under the firm's optimal product strategy offers insights for policy interventions such as consumer educational initiatives and financial incentives to foster innovation.

5 - From One-Size-Fits-All to Equitable Stratification: An Explainable Framework for Personalized Post-Cancer Surveillance

Reza Skandari, Imperial Business School, London, United Kingdom, Niloofar Zamani, Narges Mohammadi

Personalized medicine promises to improve health outcomes and reduce disparities by tailoring treatment guidelines to individual characteristics. However, its implementation is often hindered by complexity and scalability. We propose an *explainable clustering framework* to optimize the intensity and structure of population stratification, aiming to minimize health inequalities while maintaining the practicality of clinical guidelines. The population is partitioned into a manageable number of clusters, each assigned a single treatment policy—striking a balance between simplicity and equity. To ensure interpretability, the clustering is constrained to follow a decision tree structure.

We apply this method to the post-treatment surveillance problem for head and neck cancer patients, where the objective is to develop a cost-effective strategy that limits cancer recurrence detection delays to within six months. We define health-care outcome inequality as the deviation in expected detection delay from this target across subpopulations. Our results reveal diminishing returns in inequality reduction as the number of clusters increases, with the largest gains achieved by stratifying into just a few groups. Notably, the features selected in the optimal decision tree often diverge from clinical intuition, suggesting a misalignment between clinical prognosis and operational decision-making. Furthermore, we observe that as the decision tree deepens—allowing more clusters—earlier feature splits may be revised or replaced, highlighting the dynamic nature of optimal stratification as model complexity increases.

6 - Individual-Level Decision-Making for Self-Protection and Its Collective Impact on Disease Mitigation

Geonsik Yu, Purdue University, West Lafayette, IN, United States, Michael Garee, Mario Ventresca, Yuehwern Yih

Protecting public health from infectious diseases requires collective action, as individual behaviors—such as vaccination and mask-wearing—directly influence disease dynamics. During the COVID-19 pandemic, unexpected public responses often undermined the effectiveness of interventions, highlighting the need to understand behavioral patterns and motivations to design more effective strategies.

This study presents an agent-based simulation model that captures how individuals adjust self-protective behaviors based on their own individual characteristics such as evolving opinions about disease risk and how these decisions interact with external factors, such as public health interventions, to shape collective outcomes. To enhance realism, multiple datasets were integrated to generate artificial populations and calibrate individual behavior against observed patterns.

Simulation experiments were conducted under varying conditions, including optional non-pharmaceutical interventions (NPIs), heterogeneous responses to pro- and anti-intervention messaging, and alternative vaccine eligibility policies. Results reveal how decision-making patterns differ across demographic groups and how individual behaviors interact with interventions to influence disease spread.

The model explains real-world phenomena, such as the decline in mask-wearing following the first vaccination phase and disproportionately high infection rates among economically disadvantaged populations. It also identifies subpopulations likely to experience greater pressure under NPI mandates and shows how vaccination strategies—when accounting for behavioral dynamics—can significantly improve disease control and reduce disparities.

By offering insights into behavioral responses that are difficult to observe directly during real-world epidemics, this work underscores the importance of integrating behavioral dynamics into epidemic modeling and response planning to enhance the effectiveness and equity of public health interventions.

TC42

Building B Level 3 B302

Deep Learning for Medical Decision Making

Invited Session

Health Applications Society

Chair: Pooyan Kazemian, Case Western Reserve University, Cleveland, OH, United States

Co-Chair: Anjolaoluwa Popoola, Georgia Institute of Technology, Atlanta, GA, United States

1 - Knowledge-Informed Deep Learning for Periapical Lesion Diagnosis in Cone Beam Computed Tomography Scans

Yeonju Lee, Georgia Institute of Technology, Atlanta, GA, United States, Rui Qi Chen, Hao Yan, Muralidhar Mupparapu, Fleming Lure, Frank C. Setzer, Jing Li

Periapical lesions are one of the most prevalent oral pathologies, originating from the tissues surrounding the apex of a tooth. Timely diagnosis of these lesions is critical for improving clinical outcome. Delayed intervention can allow a minor condition to progress, often necessitating surgical intervention, and potentially resulting in tooth extraction. Dental cone beam computed tomography (CBCT) offers three-dimensional images of oral cavity, including oral structures (i.e., teeth, bones, restorative materials, and periapical lesions) in a single scan. CBCT is widely regarded as a powerful tool for diagnosing periapical lesions. However, due to their small sizes and similar intensity to surrounding structures, even experienced clinicians may face challenges in accurate detection and diagnosis. Leveraging Artificial Intelligence (AI) to fully automate the diagnosis process could tremendously improve the quality and efficiency of dental healthcare. The main hurdle in this advancement is reducing AI's reliance on a large quantity of manually labeled images to train robust, accurate, and generalizable algorithms. To tackle this challenge, we propose a knowledge-informed deep learning method designed to diagnose periapical lesion using CBCT. The uniqueness of our method is its capability of integrating qualitative oral-anatomical knowledge into the deep learning design. Building upon this, we further explore differentiating periapical lesions into subtypes (i.e., granuloma and cyst) to facilitate more appropriate treatment planning. We validate our method on a real-world dataset, demonstrating its potential for clinical application.

2 - Omgpt: a Sequence Modeling Framework for Data-Driven Operational Decision Making

Hanzhao Wang, Imperial College London, London, United Kingdom, Guanting Chen, Xiaocheng Li, Kalyan Talluri

We build a Generative Pre-trained Transformer (GPT) model from scratch to solve sequential decision making tasks arising in contexts of operations research and management science which we call OMGPT. We first propose a general sequence modeling framework to cover several operational decision making tasks as special cases, such as dynamic pricing, inventory management, resource allocation, and queueing control. Under the framework, all these tasks can be viewed as a sequential prediction problem where the goal is to predict the optimal future action given all the historical information. Then we train a transformer-based neural network model (OMGPT) as a natural and powerful architecture for sequential modeling. This marks a paradigm shift compared to the existing methods for these OR/OM tasks in that (i) the OMGPT model can take advantage of the huge amount of pre-trained data; (ii) when tackling these problems, OMGPT does not assume any analytical model structure and enables a direct and rich mapping from the history to the future actions. Either of these two aspects, to the best of our knowledge, is not achieved by any existing method. We establish a Bayesian perspective to theoretically understand the working mechanism of the OMGPT on these tasks, which relates its performance with the pre-training task diversity and the divergence between the testing task and pre-training tasks. Numerically, we observe a surprising performance of the proposed model across all the above tasks.

3 - Learning Clinically Interpretable Lab-Ordering Policies from Irregular ICU Data via Continuous-TIME Clinical Decision Transformer

Nasim Sadat Mousavi, University of Arkansas, Fayetteville, AR, United States, Shengfan Zhang

Routine laboratory testing is an essential component of monitoring critically ill patients; however, previous studies demonstrate that approximately 20–40% of tests ordered in intensive care units are redundant, providing minimal diagnostic information, rarely altering management decisions, and causing unnecessary costs. In this paper, we present a Continuous-Time Clinical Decision Transformer (CT-CDT), an offline reinforcement learning framework designed to model irregular, event-driven clinical data. The model can recommend which lab tests to order and when, conditioned on a return-to-go objective capturing information gain from abnormal values and transitions. We conducted goal-conditioned sequencing, which generated a subsequence of treatment history with prepended future goal state, and our state representation integrates heterogeneous EHR modalities—lab tests, vital signs, prescriptions, procedures, and fluid inputs—on an irregular event grid. For contextual embedding over admissions, we adopted a GPT-based transformer architecture with an attention mask and column embedding. This combination of temporal context modeling and missingness-aware encoding enables robust, clinically interpretable policy learning from sparse, heterogeneous measurements, offering clear advantages over methods that ignore timing or treat missing data as noise. We evaluate CT-CDT on preprocessed MIMIC-IV CKD cohorts against state-of-the-art offline RL baselines—Behavior Cloning, Dueling-DQN. Our findings demonstrate that transformer-based, goal-conditioned sequence models can learn clinically sensible lab-ordering policies from offline data while explicitly modeling irregularity and missingness, offering a promising foundation for clinician-in-the-loop decision support in ICU settings.

4 - Personalized Prediction and Control of Blood Glucose Levels Using Dynamic System Modeling

Anjolaoluwa Popoola, Georgia Institute of Technology, Atlanta, GA, United States, Pooyan Kazemian, Kamran Paynabar

Patients with diabetes in the ICU typically rely on two methods to monitor their blood glucose levels: blood tests analyzed in the laboratory and fingerstick tests. While fingerstick tests offer a cost-effective and convenient approach, they sacrifice accuracy, providing only approximate measurements. We propose a dynamic system model combining deep learning models and Kalman filtering to accurately predict blood glucose trajectory and assist with personalized glucose control over time. The resulting model, called a deep Kalman filter (Deep-KF) model, can be applied in the ICU and other healthcare settings. We validate our model using randomized real-world glucose and insulin clinical data. Finally, we demonstrate how our model can deliver precise blood glucose forecasts and insulin delivery recommendations and be used as a decision-support tool for healthcare providers, hence making glucose monitoring more efficient and patient-centric.

5 - Verifiable Artificial Intelligence for Medical Decision Making: Application to Skin Cancer Diagnosis

Pooyan Kazemian, Case Western Reserve University, Cleveland, OH, United States, Hank Korth, Manoj Malhotra

Proprietary Artificial intelligence (AI) medical tools promise to improve healthcare operations and medical decision making. Most of these tools are offered through cloud-based environments, requiring healthcare providers to submit patient data and trust the authenticity of the returned results. Currently, there is no practical way for healthcare providers to verify that the outputs were generated by the authentic AI model. This lack of verifiability introduces significant risks.

We present a trustless AI verification framework that cryptographically verifies the authenticity of AI model predictions without revealing patient data or proprietary model parameters. Leveraging zero-knowledge proofs, our framework attaches a succinct validity proof to each model inference, which will be verified on a permissionless blockchain (e.g., Ethereum) using a smart contract. Because sensitive patient data and model parameters remain hashed commitments, data privacy and intellectual property are simultaneously preserved.

We empirically demonstrate the approach with a convolutional neural network trained on labeled dermoscopic images to triage malignant versus benign skin lesions. The proposed AI verification framework eliminates trust assumptions between vendors (model providers) and clinicians (users), mitigates model tampering, and provides cryptographic auditability that demonstrates regulatory compliance. Beyond dermatology, our framework generalizes to any AI tool offered remotely where verifiable and privacy-preserving decision support is required.

TC43

Building B Level 3 B303

Designing Health and Wellness Interventions Based on Social Comparisons

Invited Session

MSOM: Healthcare

Chair: Steven Shechter, University of British Columbia, Vancouver, BC, Canada

Co-Chair: Shanshan Luo, UBC Sauder School of Business, Vancouver, BC, Canada

1 - Transparent or Not? Optimal Performance Feedback to Service Users

Guillaume Roels, INSEAD, Fontainebleau, France, Lin Chen

In many gamified services, in particular in e-health (e.g., Fitbit, Strava, sleep apps), users receive individual performance feedback upon service completion, which shapes their perception of goal achievement (through prospect theory) and relative status (when users are ahead-seeking or behind-averse). This feedback can vary in degrees of transparency, from disclosing individual scores to revealing only ranges of scores (e.g., top 5%). How transparent should service providers be in their individual performance feedback to maximize the utility users derive from their service? We employ a Bayesian persuasion framework to determine the optimal information disclosure policy.

2 - Empowering Chronic Disease Self-Management: Leveraging Social Comparison and Wellness Rewards Through Insurance Programs

Wenlin Chen, University of Electronic Science and Technology of China, Chengdu, China, People's Republic of, Jason Nguyen, Wenjie Huang, Chung-Li Tseng

The effective management of insurance for chronic diseases depends critically on patients' adherence to self-managed therapeutic regimens. Such adherence not only improves health outcomes but also reduces costs for insurers. Facing escalating treatment costs, insurers have increasingly adopted interventions aimed at encouraging subscribers to commit to managing their chronic health conditions, despite existing adherence barriers. In this study, we analytically examine how an insurer can strategically combine two prevalent interventions—a wellness reward and facilitating subscribers' social comparison—in chronic disease insurance management. We compare the insurer's optimal intervention strategies by evaluating resultant profits, insurance plan accessibility, and subscribers' health outcomes. Our findings highlight both synergistic and potentially conflicting interactions between wellness rewards and distinct forms of social comparison (upward and downward). These insights offer valuable managerial implications for insurers, healthcare providers, and policymakers, enabling informed decisions toward mutually beneficial outcomes.

3 - Optimizing Peer Matching for Performance Improvement

Steven Shechter, University of British Columbia, Vancouver, BC, Canada, Shanshan Luo

Motivation plays a key role in driving performance, yet maintaining it over time remains a challenge in domains such as fitness tracking and digital learning platforms. Research suggests that individuals are more motivated when paired with slightly stronger peers, but motivation declines when the performance gap becomes too large. Building on this principle, we develop a motivation effect model to optimize peer matching based on baseline performance levels. Our goal is to maximize motivated effort and sustained performance through structured matching. Using real-world fitness data, we evaluate the optimal matching strategy against other commonly used pairing methods. Our findings provide insights into how structured peer matching can be implemented in fitness programs to improve engagement and performance consistency.

4 - Smartphone Against Smartphone: An IT-Enabled Social Network Activation for Reducing Smartphone Overuse

Shaokang Yang, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, Idris Adjerd

This paper explores smartphone addiction through the lens of social network analysis. Using data from multiple longitudinal studies that track individual relationships and smartphone interactions, we provide empirical evidence that activating specific parts of one's social network affects smartphone usage patterns. Our results highlight the potential of strategic network interventions in managing digital addiction.

TC44

Building B Level 3 B304

Digital Innovation and Emerging Finance/Operations Issues

Invited Session

MSOM: iForm

Chair: Gangshu Cai, Santa Clara University, Santa Clara, CA, United States

Co-Chair: Xue Duan, N/A

1 - Selling to Cash-Constrained Nanostore: Supplier's Revisit Schedule and Trade Credit

Xin Geng, University of Miami, Coral Gables, FL, United States, Zheyu Jiang, Harihara Natarajan, Nan Yang

Nanostores (mom-and-pop stores, micro-retailers, etc.) are usually family-owned small retailers whose inventory replenishment is initiated by supplier and are usually cash-constrained. In addition, the nanostore must consider the family side cash need, which gives rise to a liquidity risk of cash shortfall in between the supplier's two visits. To save logistics cost and help the nanostore make more sales, the supplier can implement adaptive revisit schedule based on the store's cash level at each visit. Moreover, the supplier can also offer trade credit to the nanostore. We take a game-theoretic modeling approach to study the implications of the adaptive revisit interval and trade credit on the supplier's and the nanostore's operational decisions and financial performance.

2 - Selling Smarter: How Digital Tools and Risk-Sharing Shape Agricultural Innovation Adoption

Heng Chen, Kansas State University, Manhattan, KS, United States, Lingxiu Dong, Jie Ning, Ying Zhang

Adopting agricultural innovations such as new seed varieties or agrochemicals is essential for increasing productivity amid rising climate volatility. Farmers, however, often hesitate to adopt due to unfamiliarity with new agronomic practices and exposure to weather-related risks. In response, agribusinesses have begun bundling these innovations with digital decision-support platforms and weather risk-sharing programs to encourage adoption. Yet, the effectiveness and the strategic relationship between these two mechanisms - whether they act as complements or substitutes - remain unclear. We develop an analytical framework to evaluate their impacts on adoption, farmer welfare, agribusiness profits, and total crop yield in a setting where an agribusiness sells to a market of loss-averse farmers. We identify conditions where bundling these tools or offering each in isolation is most effective. Our findings offer insights for agricultural innovation deployment strategies across different farmer populations and cost environments.

3 - Impacts of Smart Vending Machines in Different Urban Settings: Tensor Completion with Spatiotemporal Data

Mingrui Zhang, University of Denver - Daniels College of Business, Denver, CO, United States, Wanning Chen

We investigate the treatment effect of deploying smart vending machines (treatment) compared to traditional vending machines (control) across various urban settings. Based on our results, we discuss potential changes in consumer purchasing behavior due to the introduction of technologically enhanced smart vending machines across diverse settings. The dataset we use spans 25 cities, 20 distinct urban settings, and 81 weeks. We employ tensor completion algorithms to predict over 70% missing sales outcomes in both control and treatment groups. Benchmarked against traditional regression models and matrix-based approaches, the tensor completion methods demonstrate superior predictive accuracy. The predictions for unobserved and non-existent region-setting-time combinations shed light on how to strategically plan and optimize smart vending machine deployments. More broadly, our analysis underscores the value of leveraging spectral algorithms such as tensor completion to conduct causal analysis on high-dimensional spatiotemporal data, empowering more informed decision-making.

TC45

Building B Level 3 B305

Challenges and Innovations in Sustainable and Socially Responsible Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Xavier Warnes, Indiana University, Bloomington, IN, United States

1 - Beyond Representation: Gender Diversity and Water Point Functionality in Rural Africa

Chengcheng Zhai, Univeristy of Notre Dame, South Bend, IN, United States, Kaitlin Wowak

In many rural areas of sub-Saharan Africa, community-led Water User Committees (WUCs) are responsible for maintaining village water points. However, service disruptions remain widespread. While leadership and group composition are known to influence organizational performance, little is understood about how gender dynamics within these informal governance structures affect access to clean water. We collect field data from 416 rural water points in Malawi, including detailed information on WUC composition and governance. This is matched with longitudinal functionality records from 2016 to 2024. Our outcomes measure the percentage of days each water point is fully working, in minor breakdown, or in major breakdown status. To estimate the impact of chair gender and WUC gender diversity on functionality, we use fractional logit models, propensity score matching, exact matching, and two-stage residual inclusion methods. We find that water points with female chairs perform significantly better, experiencing more fully working days and fewer minor breakdown days, than those with male chairs. In contrast, greater gender diversity within the WUC is associated with lower functionality, particularly when the chair is male. These findings are robust across model specifications and subsamples, addressing potential endogeneity concerns.

Our results suggest that simply increasing female representation on WUCs is not sufficient: placing women in leadership roles, particularly as chair, is key to improving water point functionality. NGOs and local governments should prioritize gender-inclusive leadership development and support women in decision-making positions.

2 - Multiplicity in Product Expiration Dates and Food Waste in Grocery Retail Stores

Ashish Kabra, University of Maryland-College Park, College Park, MD, United States, Nitish Jain, Varun Karamshetty

Expiration waste (EW) occurs when perishable grocery items are unsold past their expiration dates. A key contributor to EW is multiple-dates-led expiration waste (MDEW), which arises when products with varying expiration dates are simultaneously available. Consumers tend to select items with later dates, increasing the likelihood of earlier-dated items expiring. While retailers attempt to manage this through markdowns or inventory rotation, many lack visibility into MDEW's scale.

In collaboration with a grocery retailer, we assemble a multi-category-multi-store dataset (~15.3 million transactions) involving items with 3–14 days of shelf life. On average, EW accounted for 23% of sales. To quantify MDEW's share, we propose a novel and easy to implement methodology for computing its lower and upper bounds. In our retailer's context, the MDEW's lower and upper bounds equal 25% and 52% on average of the generated EW, respectively.

This study offers the first large-scale evidence of MDEW's impact and provides a practical tool for retailers to quantify and manage it, enabling more effective strategies to reduce waste and improve operations.

3 - Smarter Products, Shorter Lifespans: The Environmental Cost of Embedded Software

Sytske Wijnsma, UC Berkeley Haas, Berkeley, CA, United States, Beril Toktay, Dominique Lauga

Embedded software, due to its rapid development, has been shortening hardware lifespans, increasing e-waste. This paper examines how supply chains address obsolescence through legacy software support or hardware overbuilding. We analyze the impact of integrated vs. decentralized models and assess how cybersecurity regulations shift responsibility, influencing firm incentives, environmental outcomes, and consumer risk exposure.

4 - The Moto-Taxi Industry in Rwanda: Transition to Electric Motors

Maqbool Dada, Johns Hopkins University, Baltimore, MD, United States, Ali Fattahi, Yangfang Helen Zhou

In Rwanda, motorbikes (motos) are the pre-dominant vehicle for taxi fleets. Starting in 2019, the government announced a 5-year plan to phase out the use of internal combustion engines. Now, taxi services are operated by drivers using motorbikes. The key feature of taxi-motos is that they use portable rechargeable batteries that must be swapped periodically at charging stations as stored energy is depleted. Some e-bikes come with a single battery while others are equipped with offers two smaller swappable batteries. Since the driver has to visit a swapping center, as charge is depleted the operating range for accepting fares decreases. We build and analyze an integrated stochastic optimization model that derives policies that determines when to swap batteries; the timing is mediated by a look-ahead fare acceptance policy which depends on the distance from the current location and the passenger destination. A case study seeded from industry data is used to illustrate model application.

TC46

Building B Level 3 B306

Homeland Security and Game Theory

Invited Session

Decision Analysis Society

Chair: Jun Zhuang, University at Buffalo, Buffalo, NY, 14228, United States

1 - Understanding Active Shooter Dynamics Through Adversarial Inverse Reinforcement Learning: from Human-Subject Data to Context-Aware Multi-Agent Policies**Mohamad Dehghan Bonari, Department of Industrial and Systems Engineering, Mississippi State University, Starkville, MS, United States, Daniel Carruth , Haifeng Wang, Jun Zhuang, Randy Buchanan , Mohammad Marufuzzaman**

The growing frequency of active shooter incidents has become a major global concern, many of which occur in indoor public spaces such as schools, workplaces, and hospitals. Indoor active shooter incidents present complex multi-agent dynamics involving attackers, civilians, and law enforcement responders in confined environments. This interplay of human-human and human-building interactions, along with the uncertainty of individual behavior under extreme situations, demands a comprehensive understanding of how spatial configurations and real-time human decisions influence incident outcomes. Accordingly, we propose a scenario-based Adversarial Inverse Reinforcement Learning (AIRL) framework to guide the behaviors of agents (e.g., civilians, shooters, and law enforcement) under various, near-real-world active shooter scenarios in multi-story building settings. Our framework also captures the leader-follower behavior among civilians, where certain individuals tend to follow better-informed occupants. Moreover, we conducted a human-subject active shooter simulation study, collecting trajectory data from 40 participants through video recordings, along with their decision-making reasoning via pre- and post-trial surveys. Our proposed AIRL model benefits from trajectory data to learn the objective of each agent under different circumstances. Finally, multiple realistic scenarios are designed to evaluate the performance of the model by changing key factors in building layout configurations (e.g., entrances, hideouts, floorplans, stairs), shooter characteristics (i.e., number of assailants, weapon types, entry points), civilian distributions (i.e., the density and initial position of individuals in the buildings), and law enforcement strategies (response tactics and entry points).

2 - Evolutionary Game Analysis on Phishing Scams Using Attacker and Defender Model**Naraphorn PAOPRASERT, Faculty of Engineering, Kasetsart University, Bangkok, Thailand, Thinzar Aung, Pornthep Anussornnitarn**

This study aims to model and analyze the strategic interactions between phishing call center scammers and authorities using game theory. By identifying equilibrium stability conditions, it provides insights into optimizing cybersecurity defense strategies. The model was developed to represent the attacker (phishing call center scammers) - defender (authorities) interactions, incorporating key parameters such as costs, rewards, penalties, and success rates. The stability, primarily focusing on four pure strategy equilibrium points, was analyzed using the trace-determinant method, deriving inequalities from the eigenvalues of the Jacobian matrix. Stability conditions were examined under both general and specific assumptions to reflect realistic cybersecurity scenarios. The study identified scenarios where individual-target defense is more effective against low-cost attacks, while investing in overarching defense is preferable for countering high-cost attacks. The stability analysis revealed key factors influencing long-term strategic balance between attackers and defenders, including costs, arrest and defense rates, and successful attack rates. The findings offer important practical implications for cybersecurity policy, law enforcement, and corporate security measures. As phishing attacks become increasingly AI-driven and sophisticated, traditional, individual-target defense methods are proving less effective. This shift results in higher success rates for the attackers, highlighting the need for a transition toward overarching defense strategies, including AI-driven detection, legal enforcement such as blockage of necessary infrastructure, and public awareness campaigns to enhance cybersecurity resilience.

3 - Preferences for Security Complexity in a Constant Expected Value Environment: Evidence from a Multi-Layer Target Selection Task**Kevin Kapadia, University of Southern California, Los Angeles, CA, United States, Ian Unson, Jun Zhuang, Richard John**

Understanding how individuals evaluate structural complexity in decision-making is critical for modeling attacker behavior. In this study, 366 participants completed 12 decision scenarios in which they chose among three targets protected by one, two, or three sequential security layers. All targets were equated in expected value (EV), isolating complexity as the defining decision feature. Participants were randomly assigned to one of four between-subjects conditions crossing security layer dependency (dependent vs. independent) and incentive framing (incentivized vs. non-incentivized). Target selections were modeled using mixed-effects binary and ordinal logistic regressions. Participants were significantly more likely to choose multi-layered targets over 1-layer targets ($OR = 4.63, p < .001$). In the full ordinal model including external measures, greater boldness ($OR = 1.59, p = .009$), lower meanness ($OR = 0.61, p = .014$), and lower financial risk-taking ($OR = 0.67, p = .014$) predicted preference for more complex options. Importantly, both linear and quadratic effects of scenario EV were significant, suggesting that preferences for complex targets followed a non-linear trend across different EV configurations. Additionally, a significant interaction between information scheme and scenario EV revealed that participants in the independent condition were less likely to choose multi-layered targets in higher EV contexts. Incentive framing did not predict target selection. These findings suggest that, even when EV is controlled, perceived structural complexity carries subjective value influenced by both task framing and individual traits. Implications are discussed for security modeling, complexity preference, and personality-based prediction of adversarial behavior.

4 - Homeland Security Games in the Arctic**Jun Zhuang, University at Buffalo, Buffalo, NY, United States**

We will discuss the recent developments of homeland security games in the arctic, including foreign partners, federal, state, tribal, and local communities.

5 - Strategic Adaptation and Deception Detection Under Uncertainty in Multi-Round Attacker-Defender Games with LLMs**Sabiha Unal Eyi, University at Buffalo, Buffalo, NY, United States**

This study investigates how large language models (LLMs) perform strategic decision-making under uncertainty in attacker-defender games involving asymmetric information and potentially deceptive communication. The analysis begins with a replication of the 2x4 experimental designs introduced, where an attacker selects between two targets based on the defender's value signals and public security announcements. Across eight single-shot scenarios, LLMs were prompted to make a decision and justify their reasoning. Results show that the model predominantly targets the higher-value location and demonstrates a mix of trust and skepticism toward the defender's messages while consistently maintaining an aggressive posture by never aborting the mission.

Building on this foundation, we extend the framework to a multi-stage game setting, where the LLM plays multiple rounds against a dynamic defender. In this version, the model receives feedback on the outcome of each round and must update its beliefs accordingly. The multi-round analysis reveals that the LLM adapts its strategy over time, recognizes deceptive patterns in repeated announcements, and adjusts its target preferences based on prior outcomes and defender behavior. Thematic coding of justifications indicates emerging evidence of belief revision, pattern recognition, and rudimentary deception detection.

This work contributes to the understanding of LLMs as decision-making agents in adversarial environments and highlights their potential for strategic adaptation, especially when exposed to iterative feedback. It bridges behavioral game theory and AI by demonstrating how language models can develop context-sensitive responses in environments characterized by uncertainty and strategic signaling.

TC47

Building B Level 3 B307

Interior Point Methods

Contributed Session

Contributed

Chair: Leandro Farias Maia, Oregon State University

1 - Breaking the Barrier of Self-Concordant Barriers: Faster Interior Point Methods for M-Matrices

Adrian Vladu, CNRS, Paris, France

We study two fundamental optimization problems: (1) scaling a symmetric positive definite matrix by a positive diagonal matrix so that the resulting matrix has row and column sums equal to 1; and (2) minimizing a quadratic function subject to hard non-negativity constraints. Both problems lend themselves to efficient algorithms based on interior point methods (IPMs). For general instances, standard self-concordance theory places a limit on the iteration complexity of these methods at $O(\sim(n^{1/2}))$, where n denotes the matrix dimension. We show via an amortized analysis that, when the input matrix is an M-matrix, an IPM with adaptive step sizes solves both problems in only $O(\sim(n^{1/3}))$ iterations. As a corollary, using fast Laplacian solvers, we obtain an ℓ_2 flow diffusion algorithm with depth $O(\sim(n^{1/3}))$ and work $O(\sim(n^{1/3} \cdot \text{nnz}))$. This result marks a significant instance in which a standard log-barrier IPM permits provably fewer than $\Theta(n^{1/2})$ iterations.

(Appears in STOC 2025, <https://arxiv.org/abs/2504.20619>)

2 - An Interior Point Method Using Mollified Indicators

Daniel Turizo, Georgia Institute of Technology, Atlanta, GA, United States, Daniel Molzahn

We propose a novel interior point method using a new type of barrier. The barrier is derived by approximating the indicator function of the positive orthant via a sequence of functions called mollified indicators. These functions are infinitely differentiable, possess the properties of a barrier (they tend to infinity at zero), and have compact and shrinking support (unlike the logarithmic barriers). We develop a Newton iteration using the mollified indicator as barrier, and show that inequality constraints lying outside the barrier's support can be removed to reduce the size of the Jacobian linear system. The proposed method behaves like an active set method, where "inactive" constraints are those outside the barrier's support. On the other hand, this method is of the interior point type, so it does not possess any of the combinatorial difficulties inherent to active set methods. We performed numerical experiments comparing our method to the standard interior point algorithm with logarithmic barrier. The results show that the amortized cost of factoring the Jacobian is significantly reduced in our method.

3 - A Primal Interpretation of the Bound Flipping Ratio Test

Clifford Blakestad, SimpleRose, St Louis, MO, United States

The bound flipping ratio test (BFRT) is a powerful tool in optimizing the dual simplex method when solving linear programs. With recent advances in the theory of General Symmetric linear programs (GSLP), we can recontextualize the BFRT as a primal program phenomena. In this talk, we will discuss the relationship of the BFRT to the Restricted Ratio Test, a tool provided by using the restricted variables afforded by a GSLP. In doing so, we bridge a theoretical and computational gap between the dual and primal simplex algorithms.

4 - An Adaptive Proximal Admm for Nonconvex Composite Programs

Leandro Farias Maia, Oregon State University, Corvallis, OR, United States, David Gutman, Renato D.C. Monteiro, Gilson Silva

We developed an adaptive Proximal Alternating Direction Method of Multipliers (P-ADMM) for solving linearly-constrained, weakly convex, composite optimization problems. This method is adaptive to all problem parameters, including smoothness and weak convexity constants. Without any rank assumptions on the constraint matrices, it is shown that the adaptive P-ADMM obtains an approximate first-order stationary point of the constrained problem in a number of iterations that matches the state-of-the-art complexity for the class of P-ADMMs.

TC48

Building B Level 3 B308

Behavioral Decision Analysis I

Invited Session

Decision Analysis Society

Chair: Yucheng Su, IE Business School, MADRID, Spain

1 - Newsvendor Decisions in Regime Shift Environments

Yucheng Su, IE Business School - IE University, MADRID, Spain, Canan Ulu, Sreyaa Guha, Matthias Seifert

We examine how decision makers (DMs) adjust ordering decisions in response to demand signals that are indicative of demand regime shifts. We define demand regime shifts as fundamental changes in the underlying probabilistic demand distribution. Such shifts reflect the nonstationary nature of real-world product demand. We develop an analytical model, which extends the classical newsvendor framework, to predict ordering behavior in the face of impending regime shifts and empirically test the predictions using laboratory experiments. Our findings indicate that DMs tend to underreact to precise demand signals in unstable environments. This behavioral pattern enhances performance for low-margin products but diminishes it for high-margin products. We conclude by discussing implications of our findings for inventory management in settings where demand regime shifts are pronounced.

2 - Persuasions on Preventive Decisions

Lijia Tan, Eindhoven University of Technology, Eindhoven, Netherlands, Bochen Zhu, Rob Basten, Lijia Wei

Experts often advise decision-makers on preventive actions to mitigate potential losses. However, differences in how experts and decision-makers perceive costs and benefits can lead to a principal-agent problem. This study theoretically and experimentally examines how expert advice can persuade decision-makers to take preventive action. We propose two advisory policies: Bayesian and threshold. Our analytical results show that (1) the threshold policy outperforms the Bayesian policy in maximizing the expert's expected payoff when the conflict of interest between the expert and decision-maker is large; and (2) the two policies are equivalent when the conflict is sufficiently small. In laboratory experiments, decision-makers exhibited a relatively low adherence rate to expert advice, which reduced the expert's expected payoff. Notably, this low adherence persisted even when experts implemented the policies rationally. To capture this behavior, we employ the Agent Quantal Response Equilibrium (AQRE) model, which accounts for bounded rationality among decision-makers. Based on the irrationality estimated by the AQRE model, we propose an adapted threshold policy that improves adherence rates, thereby increasing the expert's expected payoff.

3 - Points Plus Money: Uncovering the Impact of Mixed Currency Redemptions

Freddy Lim, National University of Singapore, Singapore, Singapore, So Yeon Chun

In recent years, loyalty programs have rapidly expanded in scale. Loyalty points effectively serve as a new currency, and consumers often use loyalty point currency in the redemption of goods and services. Despite this redemption growth, consumers can still find it difficult to attain the points required to redeem rewards. In response, some firms allow consumers to perform mixed currency (points plus money) redemptions, instead of full point redemptions.

This study examines the impact of mixed currency redemptions on consumer behaviors and firm revenue, using data from a major loyalty program. We adopt an instrumental variables control function approach to model consumers' endogenous selection to perform mixed currency redemptions. We find that mixed currency redemptions significantly increase consumers' direct point purchases with the firm compared to full point redemptions, which increases firm revenue. Interestingly, earnings from traditional accrual channels such as credit cards remain similar between consumers using mixed currency and full point redemptions. This suggests that mixed currency redemptions deliver comparable impact as full point redemptions within traditional point earning channels, but uniquely drive higher levels of direct point purchases.

Although mixed currency redemptions may encourage consumers to drain their low point balances and become inactive, we find no evidence to support this. On the contrary, consumers who drain their balances through mixed currency redemptions remain engaged in the program. We also find that a consumer segment which is often overlooked in marketing efforts contributed significantly to the revenue increase. Finally, we consider plausible mechanisms that drive these results.

4 - Attention in Motion: Using Scroll Tracking to Measure the Dynamics of Preferential Choice

Ilkka Leppänen, Aalto University, Espoo, Finland

Understanding how preferences are formed in commonplace, ecologically relevant situations is highly valuable for researchers and practitioners. I develop scroll tracking, a process tracing method that uses responsive mobile web design and that can be used to study the cognitive process behind decision making. The method is based on recording a response curve in the pixel-millisecond coordinate space from the decision maker's scrolling behaviour. To demonstrate the value of response curve analysis for decision research, I propose different metrics that are derived from the response curve and that are rooted in other motor response and gaze dwell metrics in the literature. To validate the scroll tracking method I collect empirical data using an experimental study framed as a consumer choice problem and explore predictions made by attentional evidence accumulation models equipped with the scroll-based process tracing data. I also show that the way that the decision maker interacts in the app can be used to predict subjective valuations as well as latent individual difference measures and risk attitudes, as a complementary risky choice study shows. I then compare different drift diffusion models, stocked with the scroll tracking response metrics, in their predictive ability. All the models benefit from the richness of data that scroll tracking provides. These findings are evidence of the potential of naturalistic process tracing methods that are based on responsive designs of digital applications that can be used to covertly study latent decision processes, and should find use in various managerial and practical settings.

TC49

Building B Level 3 B309

Better Education: Management, Access, and Innovation

Invited Session

MSOM: Technology, Innovation, and Entrepreneurship

Chair: Owen Wu, Kelley School Of Business, Indiana University, Bloomington, IN, United States

Co-Chair: Ling Li, Indiana University, Bloomington, IN, United States

1 - Nearby Childcare Retains Talents: An Empirical Study of Childcare Availability and Turnover of US Knowledge Workers

Zhihan Wang, London Business School, London, United Kingdom, Damian Beil, Jun Li

Childcare shortage is a widely-cited critical factor hindering labor market participation in the US, yet the empirical evidence on the real impact of nearby childcare availability on the turnover of knowledge workers is still rare. Leveraging nationwide, large-scale job profile data of workers in US knowledge industries from 1997 to 2019 and variations of state-level childcare licensing regulations, this paper is among the first to investigate the causal impact of nearby childcare availability on individual-level job departure outcomes. Our results suggest that higher childcare availability leads to lower likelihood of turnover for US knowledge workers. The effect is consistent across the major occupations, and is larger for female, early-career, and highly-educated workers. Our study highlights the critical role of proximity to childcare services in talent retention and equitable workplace, offering direct policy implications for employers and policymakers.

2 - Detecting Student Struggle for Timely Intervention in Self-Paced Online Learning

Ling Li, Kelley School of Business, Indiana University, Bloomington, IN, United States, Owen Wu, Christopher Chen, Wenchang Zhang

The rise of self-paced online learning offers flexibility and personalization for K–12 students but disrupts the real-time feedback loop that exists in traditional classrooms. In physical settings, teachers observe students' facial expressions, questions, and behavior to identify confusion or frustration and intervene accordingly. However, in online environments, such visible cues vanish. Teachers are often left reviewing activity logs after the fact—an approach that is delayed, inefficient, and hard to scale. Drawing on collaborations with an online learning platform, this work addresses two critical challenges to timely instructional support in digital settings. First, students rarely signal when they are struggling. Second, instructors lack the capacity to monitor and interpret behavioral data across large student populations. While current adaptive systems personalize content delivery, they offer limited guidance on when and how teachers should intervene. Our research bridges this gap by developing data-driven methods to detect student struggle in real time, enabling just-in-time instructional support.

3 - Breaking Enrollment Barriers for Refugee Children: Evidence from An Enrollment Support Program in Turkey

Hanwen Zhu, Cornell University, Ithaca, NY, United States, Nur Kaynar, Yao Cui

Millions of refugee children around the world face persistent barriers to enrolling and remaining in school. These challenges are driven by economic hardship, language difficulties, frequent displacement, and limited institutional capacity to effectively include and support refugee students within national education systems. In this work, we collaborate with a leading non-profit organization in Turkey to evaluate the impact of enrollment-focused support services—such as information sessions on the education system, individualized counselling, and direct assistance with school registration—on school (re)enrolment and term-end retention among refugee children. Using a combination of logit models and generalized causal forests, we estimate heterogeneous treatment effects across child and household characteristics. We find that receiving support significantly improves both enrolment and retention outcomes, with the largest gains observed among younger children (ages 5–7), those with higher Turkish proficiency, and families residing in formal housing. Effects diminish sharply after age 15. These results underscore the importance of prioritizing subgroups most responsive to intervention, and suggest that targeted, low-cost services can effectively close early-grade enrolment gaps. In contrast, older adolescents and highly vulnerable families may require more comprehensive approaches—such as financial incentives tied to school attendance or structured vocational preparation programs—to sustain engagement. Situated in Turkey, which hosts the world's largest refugee population, our study demonstrates how administrative data and machine-learning-based causal inference can be leveraged to optimize the allocation of scarce educational resources in humanitarian settings.

TC50

Building B Level 3 B310

Operations and Analytics in Innovative Marketplaces

Invited Session

MSOM: Service Operations

Chair: Wenchang Zhang, Indiana University Bloomington, Bloomington, IN, United States

1 - Omnichannel Operations in on-Demand Delivery Platform with Buy-Online-and-Pick-up-in-Store

Yu Guo, University of Science and Technology of China, Hefei, China, People's Republic of, Fei Gao, Wenchang Zhang, Liu Ming

This paper studies an emerging omnichannel on-demand economy in which consumers choose how their orders are fulfilled—either via independent couriers or in-store pickup. We consider the endogenous behaviors of both consumers and couriers, and analyze the platform's optimal pricing strategies across channels and courier wage. Motivated by concerns over reduced courier income due to buy-online-and-pick-up-in-store (BOPS), we further examine the efficacy of wage regulations on platform profit, courier welfare, and consumer surplus.

We develop a stylized model to examine how BOPS affects pricing on both the consumer and courier sides. Using data from a major meal-delivery platform in China, we estimate key model primitives and simulate the effects of BOPS and wage regulation.

Our results highlight three insights. First, the conventional belief that price and demand are positively correlated for gig services may not hold in omnichannel settings—especially when consumers are less sensitive to delivery congestion. Second, although BOPS transfers delivery cost to consumers, its price may still exceed that of

delivery. Third, BOPS reduces courier welfare, but this can be mitigated by modest wage regulation: setting a minimum wage 20% above the average increases courier welfare by 15.96%, with only a 0.35% reduction in platform profit.

We conclude that traditional gig-economy pricing strategies may be suboptimal in omnichannel platforms. While BOPS boosts platform efficiency, it risks harming workers—yet a small wage floor can restore balance with minimal cost to the platform.

2 - Platform Entry and Vendor Competition

Xiaotang Yang, University of Toronto, Toronto, ON, Canada, Arseniy Gorbushin, Ming Hu, Yun Zhou

We investigate the impact of introducing a delivery platform on vendor competition. Each vendor strategically decides whether to join the platform, employ an in-house delivery, or not offer any delivery option to maximize its profit. Our findings indicate that the platform entry either intensifies vendor competition, leading to lower vendor profits and higher demand volume, or conversely, alleviates vendor competition, leading to higher vendor profits and lower demand volume. The intensifying or alleviating role of the platform on vendor competition hinges on the competitive environment of the market in which the vendors operate. We characterize conditions under which the platform's equilibrium strategy involves lowballing the delivery service to encourage vendor participation. However, perhaps surprisingly, the seemingly appealing lowballing of the delivery service (offered by the platform to attract vendor participation) hurts vendors by encouraging them to engage in more intensified competition, ultimately diminishing their profits. Conversely, if the platform overcharges from the delivery service, the platform entry alleviates vendor competition and benefits vendors. Therefore, the platform-enabled sharing of couriers can improve delivery efficiency and lower delivery costs but may intensify vendor competition through lowballing the delivery service, placing them in a prisoner's dilemma that ultimately lowers their profits. These insights also apply to an e-commerce platform that provides storage and fulfillment services with economies of scale enabled by aggregating the business of sellers operating on the platform.

3 - Incentive-Compatible Bidding and Contract Design In Digital Display Advertising

Sami Najafi Asadolahi, Santa Clara University, Santa Clara, CA, United States, Naren Agrawal, Stephen Smith

Digital advertising today operates under split incentives: advertisers fund campaigns, but agencies—or platforms—often control bidding strategies. This paper investigates that tension and proposes a coordinated solution. We develop a rigorous decision-theoretic framework to model the long-run interaction between advertisers and intermediaries, capturing how campaign pacing, budget control, and bidding incentives shape system performance over time.

We model the advertiser's and the intermediary's optimization problems as steady-state Markov decision processes and structure their interaction as a Stackelberg game. This formulation captures the dynamics of bidding, pricing, and campaign delivery under realistic operating constraints. Our central insight is that performance-based compensation, combined with active budget limits, naturally aligns agency bidding with advertiser objectives. Under such contracts, we prove that bidding becomes incentive-compatible and cost-efficient—achieving campaign outcomes at minimum total spend.

We compare three dominant industry structures—agency-managed, advertiser-managed, and platform-managed—and show that platforms like Google Ads, when subject to CPA or ROAS constraints, can replicate first-best efficiency. We further propose contract mechanisms that restore coordination and enable surplus sharing between parties.

Numerical experiments demonstrate that these coordinated pricing mechanisms can more than double advertiser surplus relative to decentralized models. The results provide both theoretical insight and practical guidance for designing revenue-optimal, incentive-aligned contracts in digital advertising markets.

TC51

Building B Level 3 B311

Emerging Topics in Platform Operations and Retail

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Dmitry Mitrofanov, Boston College, Brighton, MA, United States

1 - Nudging Customers to Select High Stock Delivery Windows via Information Sharing

Dmitry Mitrofanov, Boston College, Brighton, MA, United States, Do Yoon Kim, Benjamin Knight

For decades, stockouts have posed a major challenge for retailers, and the rise of online platforms has not mitigated this issue, now nearing a trillion-dollar scale globally. Despite AI and other innovations, companies continue to struggle with order fulfillment. This paper examines a novel policy that nudges customers toward high-stock delivery windows to improve fulfillment rates. We evaluate the effectiveness of this policy using a large-scale field experiment with over a million users on Instacart, where customers in the treatment group were presented with the prompt 'Higher stock at this time' beneath delivery windows associated with peak supermarket inventory levels. Our cost-effective information-sharing policy increases the likelihood of customers selecting high-stock delivery windows by 8.31%, with stronger effects for users with a history of stockouts, higher variability in past delivery time slots, and larger order sizes. It also reduces item replacements by 1.77% and refunds by 2.08%, improving order fulfillment. Importantly, customer spending increases, driven by a 3.91% boost in order value and a 0.72% increase in conversion rates, while order frequency remains stable. However, we also find some negative externalities: treated users spend 19.88% more time at checkout, potentially making online shopping more time-consuming. Additionally, priority delivery selection drops by 3.68%, and delivery speed decreases by 5.8%, suggesting a trade-off between successful order fulfillment and service responsiveness. Finally, the policy encourages exploration, leading customers to visit new stores and try new items, potentially generating long-term benefits for both customers and the platform, further highlighting the power of information sharing.

2 - Fair Workweek, Fair Outcomes? the Heterogeneous Treatment Effects of Predictive Scheduling Laws

Yao Cui, Cornell University, Ithaca, NY, United States, Wee Kiat Lee, Karan Girotra

Workers in the service industry tend to have varying work schedules from week to week. Such unpredictable work schedules can negatively affect the welfare of service workers. Motivated by recent legislative efforts to implement a predictive scheduling law, where employers must give advance notice for any schedule changes, we build a game-theoretic model to analyze the effect of the law. We find that while the firm's hiring level may increase with the regulation, the contracted shift hours and social welfare can potentially decrease. We also find empirical evidence for the predictions from our model using data from the state-wide implementation of a predictive scheduling law in Oregon. These insights contribute to the broader discussions on how scheduling regulations impact labor markets and firms' hiring decisions.

3 - Scheduling with Uncertain Holding Costs and Its Application to Content Moderation

Thodoris Lykouris, Massachusetts Institute of Technology, Cambridge, MA, United States, Caner Gocmen, Deeksha Sinha, Wentao Weng

In content moderation for social media platforms, the cost of delaying the review of a content is proportional to its view trajectory, which fluctuates and is a priori unknown. Motivated by such uncertain holding costs, we consider a queueing model where job states evolve based on a Markov chain with state-dependent instantaneous holding costs. We demonstrate that in the presence of such uncertain holding costs, the two canonical algorithmic principles, instantaneous-cost ($\text{\$}\nu\text{\$}$ -rule) and expected-remaining-cost ($\text{\$}\mu/\theta\text{\$}$ -rule), are suboptimal. By viewing each job as a Markovian ski-rental problem, we develop a new index-based algorithm, "Opportunity-adjusted Remaining Cost (OaRC)", that adjusts to the opportunity of serving jobs in the future when uncertainty partly resolves. We show that the regret of OaRC scales as $\tilde{O}(L^{1.5}\sqrt{N})$, where L is the maximum length of a job's holding cost trajectory and N is the system size. This regret bound shows that OaRC achieves asymptotic optimality when the system size N scales to infinity. Moreover, its regret is independent of the state-space size, which is a desirable property when job states contain contextual information. We corroborate our results with an extensive simulation study based on two holding cost patterns (online ads and user-generated content) that arise in content moderation for social media platforms. Our simulations based on synthetic and real datasets demonstrate that OaRC consistently outperforms existing practice, which is based on the two canonical algorithmic principles.

4 - Online Matching and Market Imbalance

Benjamin Barrientos, MIT, Cambridge, MA, United States, Daniel Freund, Daniela Saban

Our work introduces the effect of supply/demand imbalances into the literature on online matching with stochastic rewards in bipartite graphs. We provide a parameterized definition that characterizes instances as over- or undersupplied (or balanced), and show that higher competitive ratios against an offline clairvoyant algorithm are achievable, for both adversarial and stochastic arrivals, when instances are more imbalanced. The competitive ratio guarantees we obtain are the best-possible for the class of delayed algorithms we focus on (such algorithms may adapt to the history of arrivals and the algorithm's own decisions, but not to the stochastic realization of each potential match).

We then explore the real-world implications of our improved competitive ratios. First, we demonstrate analytically that the improved competitive ratios under imbalanced instances is not a one-way street by showing that a platform that conducts effective supply- and demand management should incorporate the effect of imbalance on its matching performance on its supply planning in order to create imbalanced instances. Second, we empirically study the relationship between achieved competitive ratios and imbalance using the data of a volunteer matching platform.

5 - Assortment Optimization, Price Competition and Fairness

Ozge Sahin, Johns Hopkins Carey Business School, Baltimore, MD, United States, Wentao Lu, Ruxian Wang

Online retailers and department stores often serve as marketplaces for multiple sellers, with consumers depending on the platform's assortment and display decisions to explore products and make purchase choices. We examine how the platform's assortment optimization influences price competition among sellers. Our analysis shows that optimizing the assortment leads to higher equilibrium prices. We then extend the model by introducing fair market exposure constraints into the platform's decision problem and study the resulting price equilibrium.

TC52

Building B Level 3 B312

Responsible Operations: Child Labor, Mining Operations, Foster Care, Financial Impact

Invited Session

MSOM: Sustainable Operations

Chair: Ashish Kabra, Nanyang Technological University, Singapore, Singapore, Singapore

1 - Search and Matching for Adoption from Foster Care

Ludwig Dierks, University of Illinois at Chicago, Chicago, IL, United States, Vincent Slaugh, Utku Unver, Sven Seuken, Nils Olberg

To find families for the more than 100,000 children in need of adoptive placements, most United States child welfare agencies have employed a family-driven search strategy in which prospective families respond to announcements made by the agency. However, some agencies have switched to a caseworker-driven search strategy in which the caseworker directly contacts families recommended for a child. In this talk, we present our work comparing these strategies, including new empirical evidence comparing a Florida agency that switched to caseworker-driven search with statewide benchmarks.

2 - Impact of Supply Chain Structures on Child Labor Violations

Jiacong Bao, R.H. Smith School of Business, UMD, College Park, MD, United States, Fred Bao, Ashish Kabra

This study examines child labor violations in US supply chains. Our analysis reveals that extended supply chain tiers and supplier network complexity significantly increase child labor violation risks, particularly in sectors with high subcontracting. Results emphasize the need for enhanced sourcing and enforcement strategies.

3 - Do Penalties Promote Responsible Operations? Evidence from the U.S. Mining Industry

Anna Saez de Tejada Cuenca, IESE Business School, Barcelona, Spain, Vibhuti Dhingra

Financial penalties—fines levied for violations found during site inspections—are widely used to encourage responsible operations (such as compliance with environmental and labor standards or with supplier codes of conduct). However, the effectiveness of penalties in achieving this objective remains unclear. In this paper, we empirically investigate whether higher penalties lead to more responsible operations using a large, granular dataset from the U.S. mining industry. Our identification strategy exploits a policy shock that sharply increased the penalty amounts for health and safety infractions in the mines. We find that higher penalties had, at best, a modest effect on responsible operations. While there was some decrease in violations, we find no evidence of a decrease in accidents or injuries at mines that faced higher fines. We show that this null effect is driven, at least in part, due to an unanticipated and unintended consequence of higher penalties: mining companies strategically contested the higher fines in court, that led to discounted payment amounts, delay in payment times, and higher regulatory costs. By examining the mechanisms through which higher penalties (fail to) promote responsible operations, our work offers insights that can be generalized to contexts beyond mining (e.g., global supply chains).

4 - Quantifying and Ranking the Impact of Social Unsustainability in Public Firms and Their Supply Chains

Christian Blanco, Ohio State University, Columbus, OH, United States, Soh Hyun Chu, James Hill

We quantify the effects of six social sustainability issues on firm market value by classifying articles from 187 news outlets published between 1990 and 2019. While the market effects are short-lived, the severity of these issues varies depending on whether they occur within the firm or its supply chain.

5 - Financial OR Productivity Interventions? A Nuanced View into Improving Farmer Welfare and Reducing Child Labor in Commodity Supply Chains

Andreas Kilian Gernert, KLU, Hamburg, Germany, Andre Calmon, Dan Iancu, Luk Van Wassenhove

Millions of smallholder farming households that produce agricultural commodities grapple with poverty and resort to their children's labor to boost farm production. In response, governments, commodity-buying firms, micro-finance institutions, and non-profit organizations have implemented various interventions to improve household welfare and mitigate child labor, but the effectiveness of these interventions remains unclear. Drawing inspiration from cocoa supply chains, we propose a model that captures a farming household's need to secure basic subsistence, alongside its decisions on borrowing, saving, consumption, and reliance on child labor for production. With this model, we evaluate the effects of common interventions—including improved access to credit and savings, price premiums, and yield improvement programs—examining their impacts on household welfare, consumption, and child labor. Our theoretical results reveal a complex landscape: improved access to credit can either decrease or inadvertently increase child labor, depending on whether households borrow for subsistence or discretionary expenses. Enhancing access to savings invariably curtails child labor, but might also depress household consumption. Price premiums can diminish child labor use only when they are large; otherwise, they may inadvertently incentivize increased child labor. Yield improvement programs can either decrease or increase child labor, depending on the household's risk aversion and prudence. We test and confirm several of these results using empirical survey data obtained from our industry partner. Our results highlight the dual-edged nature of interventions and underscore the need for tailoring strategies to the specific operational dynamics of farming households, ensuring interventions do not inadvertently exacerbate the issues they aim to resolve.

TC53

Building B Level 3 B313

Advances in Revenue Management

Invited Session

Revenue Management

Chair: Tong Xie, Booth School of Business, Chicago, IL, United States

Co-Chair: Rene Caldentey, The University of Chicago, Chicago, IL, United States

1 - Incorporating Choice Overload into Assortment Planning

Tong Xie, The University of Chicago Booth School of Business, Chicago, IL, United States, Rene Caldentey, Srikanth Jagabathula

Consumers often experience choice overload when faced with too many product options, which can hinder decision-making and reduce sales. Traditional assortment models assume rational behavior and overlook this effect. We study how choice overload shapes purchase behavior, emphasizing its role in limiting consumers' ability to explore and switch brands over time. Using panel data from the IRI Academic Dataset, we find an inverted U-shaped relationship between assortment size and brand switching, suggesting an optimal assortment size that maximizes exploration.

To capture these dynamics, we introduce the Nested Logit Choice Overload (NL-CO) model, which adds a disutility term for large assortments, balancing the benefits of variety with the costs of overload. We formulate the retailer's optimization problem under this framework and prove it is NP-hard. We then propose a 1/2-approximation algorithm and a Fully Polynomial-Time Approximation Scheme (FPTAS) to efficiently compute near-optimal assortments.

We apply the NL-CO model to private-label planning and fair assortment planning. For private labels, assuming national brands yield uniform revenue (lower than private labels) and start with larger market shares, the optimal assortment can be identified by ranking products with an index representing market share and brand-switching shifts. For fair assortment, we show that as consumers become more overloaded, the platform's ability to influence product exposure weakens. Together, these applications demonstrate how the NL-CO model captures consumer behavior under overload and provides strategies for promoting private-label growth and fairness in assortment planning.

2 - Spatial Staffing

Ming Hu, University of Toronto, Toronto, ON, Canada, Yang Bo

We study the staffing problem for an on-demand vehicle-sharing platform operating over a d -dimensional service region. The platform makes a one-time capacity decision of how many vehicles to employ at the start of an infinite horizon, and dynamically controls vehicle-customer matching and routing over time. The objective is to minimize the long-run average cost, which includes vehicle operations costs and customer waiting costs. We show that the optimal staffing level consists of a nominal load (i.e., the minimum number of vehicles to ensure system stability) plus a safety buffer. This safety level depends on key system parameters, including the dimensionality d , the distributions of customer origins and destinations, and the ride-pooling capacity q (i.e., the maximum number of passengers per vehicle). Specifically, (i) when $q=1$, the safety level scales with the arrival rate raised to the power of $d/(d+1)$, a result that holds broadly across distributional assumptions and mirrors earlier findings under stylized spatial models. (ii) When $q>1$, the scaling becomes $2d/(2d+1)$, which, to our knowledge, was not discovered in the past. Importantly, we derive these results by analyzing exact vehicle routing without relying on stylized assumptions and by developing provably near-optimal policies for general spatial matching settings beyond those in prior work.

3 - An Operational Perspective to Online Content Moderation

Tingrui Shi, University of Chicago, Chicago, IL, United States, Rene Caldentey, Amy Ward

Social media platforms and their content moderation systems heavily rely on artificial intelligence (AI) and machine learning (ML) due to the vast scale of online content they are expected to manage. These technologies enable platforms to efficiently monitor and filter user-generated content, addressing challenges such as hate speech, misinformation, and other harmful material. Although AI significantly aids in moderating vast amounts of content, human reviewers remain essential for handling ambiguous cases that AI struggles to classify reliably. Additionally, humans play a crucial role in generating new labeling data to continually train these automated systems, ensuring they remain effective and up-to-date. Keeping humans in the loop is particularly important for addressing challenges arising from an ever-evolving environment, including concept drift caused by changes in social norms, language, and social media platform policies. With a finite human-reviewer capacity, there is a trade-off between (1) dedicating more resources to labeling training content, which increases the accuracy of the machine learning classifier, and (2) devoting more resources to classification tasks, where multiple reviewers aggregate results. We examine the operational problem of optimally allocating reviewers' time between content labeling (for training machine learning models) and policy enforcement to minimize long-term misclassification and delay costs.

4 - A New Estimator for Encouragement Design in Randomized Controlled Trials when the Exclusion Restriction is Violated

Cheng Lu, Washington University in St. Louis, St. Louis, MO, United States, Guangying Chen, Tat Chan, Zhengling Qi, Dennis Zhang

Encouragement design is widely used in randomized controlled trials when noncompliance in the treatment group, control group, or both is non-negligible. The standard identification strategy is to use the randomized group assignment as an instrumental variable to estimate the local average treatment effect (LATE). In many experiments, however, this instrument may violate the exclusion restriction condition, because the encouragement can directly impact the interested outcome variable. We develop a new root-n consistent estimator using the randomized group assignment to construct an instrument that relies on the heteroskedasticity of treatment intensities between groups. Our identification strategy can recover not only LATE but also the direct impact of the encouragement on outcomes. We further propose a min-max estimator for consistent nonparametric estimation of heterogeneous treatment effects. Finally, we conducted a large-scale field experiment with a social media platform to study how expanding users' social networks influences their platform usage. While ordinary least squares and standard two-stage least squares estimators report a positive effect, our estimator suggests that the effect comes solely from the encouragement. We find evidence supporting the null effect of network expansion, indicating that firms may waste resources on false positives when the exclusion restriction is violated in their field experiments.

5 - Strategic Implications of Free Cancellation Policies in Competitive Airline Markets

Surui Wang, School of Data Science, Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Zizhuo Wang

We developed a stylized model to analyze the impact of free cancellation policies on airline competition. We study a duopoly market with two airlines offering fixed fares across two booking periods: an early period with lower fares and cancellation uncertainty, and a late period with higher fares and certain travel needs. Passengers choose between booking early, considering potential cancellations, or waiting until their travel plans are confirmed. Airlines strategically determine their booking limits and whether to provide refundable tickets. Using a utility-based economic framework, we explicitly model passenger decisions based on expected utility maximization. We derive closed-form equilibrium booking limits and refund policies. Our analysis shows that in moderately competitive markets without extensive customer spillover, airlines adopt free cancellation policies as a dominant strategy, leading to higher early bookings and improved revenues. However, under intense competition with substantial passenger spillover, refundable policies may trigger aggressive early bookings that reduce overall profitability. This study provides theoretical insights into how free cancellation policies strategically impact airline revenue management and competitive interactions.

TC54

Building B Level 2 B216

Learning for Complex Operational Systems

Invited Session

Revenue Management

Chair: Chonghuan Wang, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Jingwen Tang, University of Miami, Coral Gables, FL, United States

1 - Two-Sided Pricing and Learning with Buyer Choice: a Generalized Sgd Method

Meichun Lin, Singapore Management University, Singapore, Singapore, Woonghee Huh

Motivated by online used-car platforms, we study pricing decisions for purchasing and selling multiple products in a two-sided market. While supply for each product is independently acquired from individual sellers, demand is substitutable and follows a choice model. With uncertainty from both supply and demand, a platform sequentially adjusts purchase and selling prices to maximize profit while satisfying inventory constraints. Moreover, the platform does not know in advance the supply and demand as functions of prices. The pricing decisions should account for learning of the supply and demand functions in addition to the two sides of uncertainty and potential stockouts. When the supply and demand functions are known, we find a class of fixed-price policies that achieve provably good performance. Our finding suggests that we can target a fixed-price policy rather than the state-dependent optimal policy when the supply and demand functions are unknown. We propose a novel generalization of the classical SGD method for learning the target prices on each side. Our algorithm with the generalized SGD is asymptotically optimal when the planning horizon is large.

2 - Dynamic Pricing with Infrequent Inventory Replenishments

Menglong Li, City University of Hong Kong, Kowloon Tong, Hong Kong, Boxiao Chen, David Simchi-Levi

We consider a joint pricing and inventory control problem where pricing can be adjusted more frequently than inventory ordering decisions. More specifically, the pricing decision is adjusted every period, while new inventories are ordered every epoch that consists of multiple periods. In this setting, the retailer determines the inventory level at the beginning of each epoch and solves a dynamic pricing problem within each epoch with no further replenishment opportunities. The optimal pricing and inventory control policy is solved by an intricate dynamic programming (DP), in which the inventory level is the state variable and the pricing policy is characterized by a function of the inventory level. We consider the situation where the demand-price function and the distribution of random demand noise are both unknown to the retailer, and the retailer needs to develop an online learning algorithm to learn those information and at the same time maximize total profit. We propose a learning algorithm by applying linear bandit techniques under the upper confidence bound (UCB) framework and prove that the algorithm converges through the DP recursions to approach the optimal pricing and inventory control policy under complete demand information. The theoretical lower bound for convergence rate of a learning algorithm is proved based on the multivariate Van Trees inequality coupled with some structural DP analyses, and we show that the upper bound of our algorithm's convergence rate matches the theoretical lower bound.

3 - Operational Data Analytics (Oda) Embedded Neural Network Learning for the Newsvendor Problem with General Demand

Jian Wu, City University of Hong Kong, Hong Kong, Hong Kong, Qi Feng, J Shanthikumar

We analyze the data-integrated contextual newsvendor problem with a general demand model and high dimensional covariates. Utilizing the structural properties in the decision problem, we develop an efficient Operational Data Analytics (ODA) embedded neural network solution.

4 - Closing the Gaps: Optimality of Sample Average Approximation for Data-Driven Newsvendor Problems

Shilin Yuan, Tsinghua University, Beijing, China, People's Republic of, Jiameng Lyu, Bingkun Zhou, Yuan Zhou

We study the regret performance of Sample Average Approximation (SAA) for data-driven newsvendor problems with general convex inventory costs. In literature, the optimality of SAA has not been fully established under both α -global strong convexity and (α, β) -local strong convexity (α -strongly convex within the β -neighborhood of the optimal quantity) conditions. This paper closes the gaps between regret upper and lower bounds for both conditions. Under the (α, β) -local strong convexity condition, we prove the optimal regret bound of $\Theta(\log T/\alpha + 1/(\alpha\beta))$ for SAA. This upper bound result demonstrates that the regret performance of SAA is only influenced by α and not by β in the long run, enhancing our understanding about how local properties affect the long-term regret performance of decision-making strategies. Under the α -global strong convexity condition, we demonstrate that the worst-case regret of any data-driven method is lower bounded by $\Omega(\log T/\alpha)$, which is the first lower bound result that matches the existing upper bound with respect to both parameter α and time horizon T . Along the way, we propose to analyze the SAA regret via a new gradient approximation technique, as well as a new class of smooth inverted-hat-shaped hard problem instances that might be of independent interest for the lower bounds of broader data-driven problems.

TC55

Building B Level 3 B314

Generative AI for Markets and Decision-Making

Invited Session

Revenue Management

Chair: Negin Golrezaei, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: Sourav Sahoo, MIT, Cambridge, MA, 02139, United States

1 - The Cost of Balanced Training-Data Production in An Online Data Market

Augustin Chaintreau, Columbia University, New York, NY, United States, Juba Ziani, Roland Maio

Many ethical issues in machine learning are connected to the training data. Online data markets are an important source of training data, facilitating both production and distribution. Recently, a trend has emerged of for-profit "ethical" participants in online data markets. This trend raises a fascinating question: Can online data markets sustainably and efficiently address ethical issues in the broader machine-learning economy?

In this work, we study this question in a stylized model of an online data market. We investigate the effects of intervening in the data market to achieve balanced training-data production. The model reveals the crucial role of market conditions. In small and emerging markets, an intervention can drive the data producers out of the market, so that the cost of fairness is maximal. Yet, in large and established markets, the cost of fairness can vanish (as a fraction of overall welfare) as the market grows. Our results suggest that "ethical" online data markets can be economically feasible under favorable market conditions, and motivate more models to consider the role of data production and distribution in mediating the impacts of ethical interventions.

2 - Econevals: Benchmarks and Litmus Tests for LLM Agents in Unknown Environments

Sara Fish, Harvard University, Cambridge, MA, United States, Julia Shephard, Minkai Li, Ran Shorrer, Yannai Gonczarowski

We develop benchmarks for LLM agents that act in, learn from, and strategize in unknown environments, the specifications of which the LLM agent must learn over time from deliberate exploration. Our benchmarks consist of decision-making tasks derived from key problems in economics. To forestall saturation, the benchmark tasks are synthetically generated with scalable difficulty levels. Additionally, we propose litmus tests, a new kind of quantitative measure for LLMs and LLM agents. Unlike benchmarks, litmus tests quantify differences in character, values, and tendencies of LLMs and LLM agents, by considering their behavior when faced with tradeoffs (e.g., efficiency versus equality) where there is no objectively right or wrong behavior. Overall, our benchmarks and litmus tests assess the abilities and tendencies of LLM agents in tackling complex economic problems in diverse settings spanning procurement, scheduling, task allocation, and pricing -- applications that should grow in importance as such agents are further integrated into the economy.

3 - Designing Enterprise AI Systems: Hallucination, Creativity, and Moral Hazard

Terry Taylor, UC Berkeley, Berkeley, CA, United States, Tinglong Dai

We explore the interplay between the design of Generative AI (GenAI) systems and the compensation of human agents. A key hyperparameter, temperature, governs the trade-off between creativity and reliability, such that higher temperatures enable more creative outputs but increase the risk of hallucinations. We examine how temperature interacts with agent effort in a principal-agent framework. The firm sets the GenAI system's temperature, while the agent chooses an unobservable effort level, both of which jointly determine the probability of success. We compare this decentralized system, where the agent chooses her effort, with the benchmark centralized system, where the firm dictates both temperature and effort.

We show that when the firm's design of its GenAI system is taken into account, classical moral hazard results are reversed: decentralization can distort agent effort *upward*; and as the value of a successful outcome increases, the optimal effort level can *decrease*. One might expect the optimal temperature to be monotone in the value of a successful outcome; indeed, this is the case for the centralized system. We show how this result is reversed for the decentralized system. Together, these findings offer a comprehensive framework for understanding the interplay among GenAI system parameters, incentive design, and performance.

4 - Rethinking the Role of Genai in Content Creation Ecosystem

Fan Yao, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

The advent of generative AI (GenAI) is transforming the landscape of content creation industry, enabling the rapid production of diverse, high-quality media. While GenAI reshapes online content ecosystems, it also raises concerns about market saturation and the potential marginalization of human creativity. In this talk, we introduce a competition model to analyze the dynamics between authentic human creators and GenAI. Our theoretical and simulation results suggest that, despite the challenges, a stable equilibrium between human- and AI-generated content is achievable under certain conditions. This study aims to advance the conversation around competitive dynamics in online content ecosystems and offer insights into the evolving interplay between human creativity and generative technology.

5 - Tokenized Bandit for LLM Decoding and Alignment

Suho Shin, University of Maryland, College Park, MD, United States

We introduce the tokenized linear bandit (TLB) and multi-armed bandit (TMAB), variants of linear and stochastic multi-armed bandit problems inspired by LLM decoding and alignment. In these problems, at each round $t \in [T]$, a user submits a query (context), and the decision maker (DM) sequentially selects a token irrevocably from a token set. Once the sequence is complete, the DM observes a random utility from the user, whose expectation is presented by a sequence function mapping the chosen token sequence and the query to a real value. In both problems, we first show that learning is impossible without any structure on the sequence function. We introduce a natural assumption, diminishing distance with more commons (DDMC), and propose algorithms with regret $\tilde{O}(\sqrt{T})$ and $\tilde{O}(\sqrt{T^{2/3}})$ for TLB and TMAB, respectively. As a side product, we obtain an (almost) optimality of the greedy decoding for LLM decoding algorithm under DDMC, which justifies the unreasonable effectiveness of greedy decoding in several tasks. This also has an immediate application to decoding-time LLM alignment, when the misaligned utility can be represented as the frozen LLM's utility and a linearly realizable latent function. We finally validate our algorithm's performance empirically as well as verify our assumptions using synthetic and real-world datasets.

TC56

Building B Level 3 B315

Combinatorial Markets

Invited Session

Auctions and Market Design

Chair: Benjamin Lubin, Boston University, Boston, 02215, United States

Co-Chair: Robert Day, University of Connecticut, Storrs, CT, United States

1 - Iterative Vickrey Auctions via Linear Programming

Sébastien Lahaie, Google Research, New York, NY, United States, Benjamin Lubin

Building on the linear programming approach to competitive equilibrium pricing, we develop a general method for constructing iterative auctions that achieve Vickrey-Clarke-Groves (VCG) outcomes. We show how to transform a linear program characterizing competitive equilibrium prices into one that characterizes universal competitive equilibrium (UCE) prices, which elicit precisely the information needed to compute VCG payments. By applying a primal-dual algorithm to these transformed programs, we derive iterative Vickrey auctions that maintain a single price path, eliminating the overhead and incentive problems associated with multiple price paths used solely for payment calculations. We demonstrate the versatility of our method by developing a novel iterative Vickrey auction for the multi-unit setting and an iterative variant of the Product-Mix auction. The resulting auctions combine the transparency of iterative price discovery with the efficiency and incentive properties of the VCG mechanism.

2 - Risk-Aware Market Clearing for Electricity Markets

HRITHIK NAMBIAR, Brown University, Providence, RI, United States, Amy Greenwald, Wirattawut Boonbandansook, Ryan Edmonds, Sidharth Anand, Gianluca Brero, Benjamin Lubin

Building on earlier work employing stochastic optimisation in two-period electricity markets, we propose a market-clearing mechanism that prices and allocates in the first stage to better manage extreme events in the second stage.

3 - Pricing Valid Cuts for Price-Match Equilibria

Robert Day, University of Connecticut, Storrs, CT, United States, Benjamin Lubin

We use valid inequalities (cuts) of the binary integer program for winner determination in a combinatorial auction (CA) as “artificial items” that can be interpreted intuitively and priced to generate Artificial Walrasian Equilibria. While the lack of an integer programming gap is sufficient to guarantee a Walrasian equilibrium, we show that it does not guarantee a “price-match equilibrium” (PME), a refinement that we introduce, in which prices are justified by an iso-revenue outcome for any hypothetical removal of a single bidder. We prove the existence of PME for any CA and characterize their economic properties and computation. We implement minimally artificial PME rules and compare them with other prominent CA payment rules in the literature.

4 - Expressive Side Information in Combinatorial Markets

Siddharth Prasad, Carnegie Mellon University, Pittsburgh, PA, United States, Maria-Florina Balcan, Tuomas Sandholm

I will present a new framework for efficient mechanism design with *expressive side information*. Expressive side information is a new paradigm where any knowledge the mechanism designer has about the agents—modeled through constraints on the agents' type spaces—can be used to design efficient mechanisms with better pricing structures. We characterize the revenue-optimal efficient mechanism under a very general model of type spaces and design better core-selecting auctions and better budget-balanced exchanges.

Joint work with Nina Balcan and Tuomas Sandholm.

5 - Service Assignment at the United States Naval Academy

Naomi Utgoff, US Naval Academy, Annapolis, MD, United States

This paper studies service assignment at the United States Naval Academy (USNA). Service assignment is the process which assigns each member of USNA's senior class to one of several communities, each representing a career path within the Navy. This matching problem is unique to the literature in that it features hard minimums, individual rationality constraints, and the requirement that every midshipman be assigned to a community. A stable assignment is generally impossible under these requirements: we characterize a natural relaxation that we call minimal unfairness. We then introduce the individual rationality adjusted deferred acceptance algorithm (IRDA) and show that it always selects a minimally unfair assignment which respects the Navy's requirements. This mechanism relies on the construction of a series of caps which combine the minimums, maximums, and individual rationality constraints. These caps thus determine when a community must reject a midshipman, as well as which midshipman should be rejected. Unlike standard deferred acceptance in which the rejecting community simply rejects its least favorite proposal(s) above its ceiling, IRDA identifies a rejectable set of midshipmen, and forces the rejection of a midshipman in that set. We also discuss two alternative mechanisms, generalized serial dictatorship (GSD) and eligibility type capped deferred acceptance (ECDA), and contrast their properties with IRDA.

TC57

Building B Level 3 B316

Selected Auctions and Market Design Papers from the EC 2025 Conference I

Award Session

Auctions and Market Design

Chair: Rad Niazadeh, Chicago Booth School of Business, Chicago, IL, United States
 Co-Chair: Jiaqi Lu, the Chinese University of Hong Kong (Shenzhen), Shenzhen, N/A
 Co-Chair: Christian Kroer, Columbia University, New York, NY, United States

1 - Competition Complexity in Multi-Item Auctions: Beyond Vcg and Regularity

Yiding Feng, Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Hedyeh Beyhaghi, Yingkai Li, Matthew Weinberg

We quantify the value of the monopoly's bargaining power in terms of competition complexity—that is, the number of additional bidders the monopoly must attract in simple auctions to match the expected revenue of the optimal mechanisms—within the setting of multi-item auctions. We show that for simple auctions that sell items separately, the competition complexity is $\Theta(n/\alpha)$ in an environment with n original bidders under the slightly stronger assumption of α -strong regularity, in contrast to the standard regularity assumption in the literature, which requires $\Omega(n \cdot \ln(m/n))$ additional bidders. This significantly reduces the value of learning the distribution to design the optimal mechanisms, especially in large markets with many items for sale. For simple auctions that sell items as a grand bundle, we establish a constant competition complexity bound in a single-agent environment when the number of items is small or when the value distribution has a monotone hazard rate. Some of our competition complexity results also hold when we compete against the first best benchmark (i.e., optimal social welfare).

2 - On the Theoretical Foundations of Data Exchange Economies

Aniket Murhekar, Northwestern University, Evanston, IL, United States, Hannaneh Akrami, Bhaskar Chaudhury, Jugal Garg

Organizations increasingly seek to share and access datasets to improve their ML models and derive insights. Despite the immense demand for quality data, data exchange and collaboration have not reached their full potential. One of the key reasons is the lack of reciprocity, where some participants perceive their contribution to others to be of higher value than what they receive in return.

3 - Inspection Game with Location-Specific Detection Capabilities

Mathieu Dahan, Georgia Institute of Technology, Atlanta, GA, United States, Bastián Bahamondes

We consider a two-person inspection game, in which a defender positions a limited number of detectors across a critical system to detect multiple attacks caused by an attacker. We assume that detection is imperfect, and each detector location is associated with a probability of detecting attacks within its set of monitored system components. The objective of the defender (resp. attacker) is to minimize (resp. maximize) the expected number of attacked components that remain undetected. To compute Nash Equilibria (NE) for this large-scale zero-sum game, we formulate a linear program with a small number of constraints, which we solve via column generation. We provide an exact mixed-integer program for the pricing problem, which entails computing a defender's pure best response, and leverage its supermodular structure to derive two efficient approaches to obtain approximate NE with theoretical guarantees: A column generation and a multiplicative weights update (MWU) algorithm with approximate best responses. To address the computational challenges posed by combinatorial attacker strategies, each iteration of our MWU algorithm requires computing a projection under the unnormalized relative entropy. We provide a closed-form solution and a linear-time algorithm for the projection problem. Our computational results in real-world gas distribution networks illustrate the performance and scalability of our solution approaches.

4 - Markets with Heterogeneous Agents: Dynamics and Survival of Bayesian vs. No-Regret Learner

Yoav Kolumbus, Cornell University, Ithaca, NY, United States, David Easley, Eva Tardos

We analyze the performance of heterogeneous learning agents in asset markets with stochastic payoffs. Agents aim to maximize the expected growth rate of their wealth but have different theories on how to learn to do this best. Our main focus is on comparing Bayesian learners and no-regret learners who compete in markets and identifying the conditions under which each approach is more effective. Bayesian learners with a finite prior that assigns positive probability to the correct model have posterior beliefs that converge exponentially fast, and such agents survive even in the presence of agents who invest according to the correct model. Bayesian learners with a continuum prior converge at a slower rate of $O((\log T)/T)$. Online learning theory provides no-regret algorithms for maximizing the log of wealth, achieving worst-case regret bounds of $O((\log T)$ without assuming a stationary stochastic process, but comparing against the best fixed investment rule. This regret, as we observe, is of the same order as that of a Bayesian learner with a continuum prior. Surprisingly, we find that even such low regret may not guarantee survival: an agent with regret $O((\log T)$, converging to the correct model at rate $O((\log T)/T)$ still vanishes when competing against an agent who invests according to the correct model, or even against a perfect Bayesian with a finite prior. On the other hand, we show that Bayesian learning is fragile, while no-regret learning requires less knowledge of the environment and is therefore more robust.

5 - Reallocating Wasted Votes in Proportional Parliamentary Elections with Thresholds

Rupert Freeman, University of Virginia, Charlottesville, VA, United States, Théo Delemazure, Jérôme Lang, Jean-François Laslier, Dominik Peters

In many proportional parliamentary elections, electoral thresholds (typically 3-5%) are used to promote stability and governability by preventing the election of parties with very small representation. However, these thresholds often result in a significant number of "wasted votes" cast for parties that fail to meet the threshold, which reduces representativeness. One proposal is to allow voters to specify replacement votes, by either indicating a second choice party or by ranking a subset of the parties, but there are several ways of deciding on the scores of the parties (and thus the composition of the parliament) given those votes. We introduce a formal model of party voting with thresholds, and compare a variety of party selection rules axiomatically, and experimentally using a dataset we collected during the 2024 European election in France. We identify three particularly attractive rules, called Direct Winners Only (DO), Single Transferable Vote (STV) and Greedy Plurality (GP). The full version of this paper is available at <https://arxiv.org/abs/2503.17>.

TC59

Building B Level 3 B318

AI and Analytics in Healthcare Services

Contributed Session

Contributed

Chair: Parisa Lotfibagha, University of Massachusetts Amherst, Amherst, MA, United States

1 - An Adaptive Large Neighborhood Search Algorithm with Customized Search Operators for Solving Home Healthcare Scheduling and Routing Problem with Skill Matching and Periodic Visits

Qu Wei, Department of Emergency Medicine and West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, 610041, China, Chengdu, China, People's Republic of, Kang Li

Home healthcare services are increasingly recognized as a crucial solution to challenges such as population aging, medical resource shortages, and growing service demands. However, efficiently scheduling healthcare services to minimize operational costs, optimize provider utilization, and ensure high-quality provider-patient matching remains a significant challenge. This study develops a comprehensive optimization model that incorporates practical constraints, including skill matching, work time limits, and periodic scheduling, to enhance service efficiency and resource allocation. Given the large-scale nature and computational complexity of the problem, traditional exact algorithms struggle to find optimal solutions within a reasonable timeframe. To address this, we propose an adaptive large neighborhood search algorithm with customized search operators designed specifically for home healthcare scheduling. The algorithm intelligently explores the solution space, improving convergence speed and solution quality. Extensive computational experiments using both synthetic and real-world data demonstrate that our approach consistently outperforms commercial solvers like Gurobi and conventional metaheuristics in both solution quality and computational efficiency, particularly in large-scale instances. A case study using real-world data from West China Hospital further validates the algorithm's practical applicability, showing significant reductions in operational and provider-related costs while enhancing service delivery. These findings offer valuable insights for healthcare managers, providing advanced decision-support tools to improve service quality, reduce expenses, and better meet the increasing demand for home healthcare services. This study contributes to healthcare operations research by presenting a scalable and efficient scheduling framework that integrates advanced optimization techniques with real-world constraints.

2 - Optimizing Chemotherapy Patient Selection and Timing for Home-Based Cancer Care

Lauren Moore, Mayo Clinic, Rochester, MN, United States, Yu-Li Huang

Timely access to chemotherapy is critical for cancer treatment outcomes. Home-based chemotherapy administered by mobile clinicians offers a safe and patient-centered alternative to in-clinic infusions, and it can reduce clinic appointment burden and patient treatment start lead time. However, not all patients are eligible or willing to receive care at home, and resources must be effectively managed. This study develops a tactical-level framework to support decision-making around the initiation timing of home-based chemotherapy under the Cancer Care Beyond Walls (CCBW) program at Mayo Clinic in Florida. A mixed-integer linear programming model (MILP) was developed to maximize utilization of home-based care resources while adhering to clinical and resource constraints. The model is adjusted for multiple policy options that reflect different levels of operational flexibility: Fixed-Term, Flexible-Term, and Recurrent. Using simulated data and weekly decision-making intervals, we evaluate the impact of each policy on patient lead time and in-clinic resource availability. Results show that the scheduling model is feasible for managing home-based patient volumes. More flexible policies may increase resource utilization and reduce average lead times, though we consider additional coordination and dynamic staffing considerations. This work anticipates the future implementation of the CCBW program, providing practical guidance for scheduling decisions and informing infrastructure planning for a sustainable practice offering.

3 - Strategic Disinformation Generation and Detection

Pengxiang Zhou, University of Southern California, Los Angeles, CA, United States, Wenxiao Yang, Yunfei (Jesse) Yao

Disinformation detection is becoming increasingly important and relevant because it is easier than ever to create and disseminate disinformation. How does detection ability affect the incentive to generate disinformation? Given the practical constraints of classification technology, how should a detector be designed? To answer these questions, this paper studies the problem where a sender strategically communicates his type (high or low) to a receiver, and a lie detector generates a noisy signal on the truthfulness of the sender's message. The receiver then infers the sender's type both through the message from the sender and through the signal from the detector. We find a non-monotonic relationship between the probability that the low-type sender is lying and the accuracy of detection. More accurate detection (a higher true-positive rate and a lower false-positive rate) increases the probability of lying when the true-positive rate is low, because of a persuasive effect. By contrast, more accurate detection decreases the probability of lying when the true-positive rate is high, due to an alarm effect. We also characterize the optimal detector design. The designer always chooses the lowest feasible false-positive rate for any true-positive rate. The possibility of false-positive alarms implies that the designer chooses an intermediate true-positive rate rather than the highest true-positive rate. Counter-intuitively, the optimal detector may raise an alarm about a smaller percentage of disinformation when its underlying classifier is better at distinguishing the sender's type.

4 - Optimizing Service Mix in Substance use Treatment: the Role of Supplemental Services in Improving Treatment Retention

Matt Baucum, Oklahoma State University, Stillwater, OK, United States, David Dreyfus

Substance use disorder (SUD) represents a significant public health challenge. Yet no research to date has investigated the impact of SUD treatment service mix on patient outcomes using nationally-representative data. In this study, we investigate whether SUD treatment retention is improved when providers offer supplemental services (i.e., 'wraparound services') to patients, and whether this effect is moderated by the barriers patients face when accessing treatment. To address this, we extract all SUD treatment facility information from the SAMHSA National Directory of Drug and Alcohol Abuse Treatment Facilities from 2017 - 2022, including location and service offering data for each facility. We link this data to metro area-level treatment records from the SAMHSA Treatment Episode Dataset, along with a state-level analysis of restrictive SUD treatment regulations. We find that SUD treatment completion rates increase when centers offer supplemental services to patients, and that this effect is strongest in areas where patients have poor access to treatment. We then investigate whether it is cost effective (from either a provider or policymaker perspective) to increase patient retention rates through supplemental service offerings. We find that the increased revenue to providers (due to longer patient stays) is generally not sufficient to cover the cost of increased supplemental service offerings. Yet we find that it is cost effective for states to increase reimbursement rates for such services, due to the positive societal spillover effects of increased SUD patient retention. This suggests a possible role for policymakers to steer SUD providers to more holistic service offerings.

5 - Anxiety vs. Trust: Emotional Factors Affecting the Medical Service Supply Chain

Li Hou, Ningbo University of Technology, Ningbo, China, People's Republic of, Mengbi Zeng

With rising living standards and an aging population, China faces a significant shortage of medical resources and an unbalanced allocation. The mismatch between high-quality resources and patient needs leads to low operational efficiency in the medical service supply chain. This study applies the Elaboration Likelihood Model (ELM) to explore the emotional drivers behind patients' irrational preferences for tertiary hospitals, which contribute to systemic inefficiencies. Moving beyond traditional analyses focused on cost and accessibility, we reveal how health anxiety and institutional distrust interact with rational cognition, exacerbating resource misallocation. Using structural equation modeling based on patient surveys, three key findings emerge: First, anxiety influences decision-making as a peripheral ELM pathway, overshadowing logical assessments of primary care suitability. Second, health literacy moderates this emotional bias; cognitively empowered patients exhibit stronger rational decision-making abilities. Third, trust-building mechanisms within primary care institutions prove unexpectedly ineffective due to deep-rooted societal stigma. This paper pioneers the use of ELM in analyzing healthcare supply chains by establishing a dual-path framework that enhances our understanding of behavioral barriers to hierarchical diagnosis reform in China. It advocates for emotion-centric interventions—specifically utilizing health education to reduce anxiety-driven choices—and redesigning community healthcare experiences to restore both visual and psychological trust. These conclusions provide policymakers with behavioral tools that complement structural reforms aimed at addressing both the “visible” resource imbalances and the “invisible” affective barriers hindering optimization within China’s medical system.

6 - A Semi-Markov Decision Process Model to Optimize Return to Clinic Decisions in Primary Care

Parisa Lotfibagha, University of Massachusetts Amherst, Amherst, MA, United States, Muge Capan, Kristen Miller, William Gallagher, Elizabeth Selden

Optimal return-to-clinic strategies for patients with type 2 diabetes (T2D) managed in primary care influence glycemic control, prevention of organ damage, healthcare utilization, and related costs. Current clinical guidelines, however, predominantly rely on Hemoglobin A1c (HgA1c) levels and recommend uniform follow-up intervals, inadequately capturing variability in patient characteristics and clinical complexities. To overcome this limitation, we developed a finite state infinite horizon Markov model to represent a subpopulation-specific return-to-clinic decision framework. The model analyzes patient health trajectories derived from comprehensive, de-identified longitudinal Electronic Health Record data (2017–2023) from two primary care settings affiliated with a large scale healthcare system in the mid-Atlantic region with 1703 unique patients diagnosed with T2D. Specifically, the framework integrates T2D related comorbidities, such as chronic cardiovascular disease, chronic kidney disease, and a smoothing function applied to HgA1c measurements to estimate unobserved glycemic values indicative of potential organ damage, alongside healthcare utilization metrics, notably emergency department visits. The results provide optimal subpopulation-specific return-to-clinic policies which deviate from the standard six to twelve months follow-up intervals recommended by the American Diabetes Association. For patients with chronic conditions and out-of-control HgA1c values the optimal policy requires more frequent intervals between subsequent primary care visits, such as one to three months. This study highlights the importance of personalized chronic disease management to achieve glycemic control. Future work will focus on the tradeoff between time interval for follow-up and the value of new clinical information and develop transition probabilities informed by additional patient characteristics to personalize return-to-clinic policies.

TC60

Building B Level 3 B319

Machine Learning and Time Series

Contributed Session

Contributed

Chair: Yufei Xu, Georgia Institute of Technology, atlanta, GA, United States

1 - Domain-General Reconstruction Through Invariant and Specific Representations

Hyewon Ryu, KAIST, Daejeon, Korea, Republic of, Hyungkwon Lee, Jong-Seok Lee

Reconstruction is a fundamental task employed across various fields, especially in unsupervised learning contexts. However, reconstruction models often struggle to generalize to unseen domains due to an inherent trade-off between preserving reconstruction fidelity and learning robust, domain-invariant representations. This limitation becomes particularly critical when data distributions shift across domains, leading to significant performance degradation. To address this challenge, we propose a novel method designed to explicitly separate domain-invariant from domain-specific information using a single encoder. By leveraging selective network connectivity, our method projects input data into two complementary latent spaces—domain-invariant and domain-specific—ensuring that domain-specific variables capture only residual information conditioned on invariant features, without reciprocal influence. We validate our approach on both synthetic and real-world multivariate time series benchmarks with induced domain shifts. Our experiments demonstrate that the proposed architecture consistently achieves lower reconstruction error and maintains stable performance on unseen domains. These results confirm that jointly modeling invariant and specific variables enables high-fidelity reconstruction while ensuring robust generalization in dynamic environments.

2 - Modernizing TIME Series Forecasting: Reuniting Statistical Structure with Deep Learning Through Transformer Innovation

Shihao Yang, ISyE Georgia Tech, Atlanta, GA, United States, Jiecheng Lu

The accelerating adoption of deep learning in time series forecasting has brought renewed attention to Transformer-based architectures. Despite their empirical success, many existing models lack interpretability and alignment with traditional statistical principles, often underperforming in multivariate or long-horizon settings. This talk presents a unifying framework that bridges classical time series models with modern deep architectures, advancing both accuracy and interpretability.

We begin by revisiting linear attention in Transformers through a Vector Autoregressive (VAR) lens, revealing structural misalignments in current deep models and introducing SAMoVAR, a VAR-aligned Transformer that restores interpretability and forecasting fidelity. Building on this foundation, we propose WAVE, a novel attention mechanism integrating ARMA dynamics via implicit moving average weights, boosting the ability to capture both local and global temporal dependencies efficiently.

Addressing scalability and context-awareness, we introduce the In-Context Time Series Predictor, which reformulates time series forecasting tasks for efficient in-context learning, eliminating the need for parameter updates while maintaining competitive performance in full, few-, and zero-shot scenarios. For complex multivariate forecasting challenges, we explore two orthogonal innovations: (1) ARM, which adapts to series-specific temporal patterns using techniques like Adaptive Univariate Effect Learning, Random Dropping, and Multi-kernel Local Smoothing; and (2) CATS, which constructs Auxiliary Time Series to enrich inter-series dependencies while minimizing model complexity.

Together, these works underscore a unified theme: bridging statistical rigor with architectural innovations in Transformers to overcome limitations in generalization, overfitting, and computational complexity.

3 - General Supervised Learning Framework for Open World Classification

Mohammad Amin Basiri, University of Oklahoma, Norman, OK, United States, Sai Krishna Theja Bhavaraju, Charles Nicholson

In many real-world classification tasks, new data may include classes not represented in the training set, posing challenges for traditional supervised learning models. This open-world classification (OWC) scenario is especially relevant in dynamic environments such as social media analytics, cybersecurity, and autonomous systems. Most existing solutions are domain-specific or tightly coupled to particular algorithms, limiting their applicability. We propose a flexible framework, adaptable to diverse data types and models, that integrates two core capabilities: (1) identifying novel data instances by measuring how different their prediction patterns are from known classes, and (2) grouping these instances into new classes based on similarity in those prediction patterns. By operating in the classifier's output probability space and decoupling detection from categorization, the framework remains flexible across domains and model types. It supports continuous model updates without human intervention, enabling real-time adaptation in dynamic settings. Experiments across diverse datasets, including text, sensor, image, and numeric data, demonstrate consistent improvements in classification performance. A case study on disaster-related Twitter data demonstrates how the framework can support real-time situational awareness by detecting emerging crisis topics, with potential applications in humanitarian coordination, resource deployment, and disaster recovery logistics.

4 - Multivariate TIME-Series Self-Supervised Learning for Domain Adaptation : Application to Tool Wear Prediction

Seonyoung Kim, Ewha Womans University, Seoul, Korea, Republic of, Soomin Lee, Dongil Kim

Tool condition monitoring is critical for manufacturing efficiency, with accurate tool state detection directly impacting process reliability. While self-supervised learning (SSL) offers promising approaches for representation learning with limited labeled data, its application to process time-series data remains underdeveloped. Our research introduces a novel SSL framework specifically designed for tool condition monitoring using sensor data collected during machining operations. The framework incorporates pretext tasks tailored to capture the temporal and contextual characteristics of multivariate time-series data, while integrating domain adaptation techniques to enhance model generalization across different operating conditions. This approach addresses the critical challenge of domain shifts that occur when process parameters change or new fault patterns emerge. Through comprehensive experiments on classification and regression tasks, we demonstrate that our method achieves superior performance compared to conventional approaches, particularly in scenarios with limited labeled data and when transferring knowledge between different machining environments, advancing the development of adaptive and intelligent tool condition monitoring systems.

5 - Detecting Temporal State Changes in Astronomical TIME Series Data Using Physics-Informed Neural Differential Equations

Seungsu Kam, Ulsan National Institute of Science and Technology (UNIST), Ulju-gun, Korea, Republic of, Yongkyung Oh, Minsu Shin, Sungil Kim

We present a novel unsupervised deep learning approach for detecting changing-state quasars (CSQSOs), which exhibit abrupt temporal transitions in their optical properties and are of growing interest in quasar research. Detecting CSQSOs is challenging due to irregular sampling, data sparsity, complex variability, and the scarcity of labeled data. To overcome these issues, we develop a physics-informed neural differential equation model that incorporates the known physical properties of quasars into its loss function, while effectively handling irregularly sampled time series. Our proposed model captures both stochastic and physical aspects of quasar behavior and performs anomaly detection based on reconstruction error. This framework enables robust CSQSO detection without labeled supervision, and its effectiveness is validated through extensive simulations using datasets derived from physically motivated quasar models.

6 - Fusing Federated Learning and Knowledge Distillation for Fair Travel Choice Prediction

Yufei Xu, Georgia Institute of Technology, atlanta, GA, United States, Zhiwei Chen, Srinivas Peeta

Accurate travel choice prediction is essential for demand modeling and transportation planning. Although Deep Learning (DL) models offer substantial potential for travel choice prediction, they often exhibit algorithmic unfairness, yielding systematically poorer predictions for disadvantaged population groups such as racial minorities and low-income populations compared to their counterparts. Existing studies that attempt to handle this issue in the transportation domain are few and often confront a steep tradeoff between accuracy and fairness due to the in-processing approaches they utilize, which sacrifice accuracy to improve fairness. This study introduces a novel DL approach that fuses federated learning and cross-client knowledge distillation to overcome this challenge. Since imbalanced data can prompt DL models to treat minority observations as noise or pay less attention, the dataset is partitioned into demographic subgroups (e.g., racial minority vs. majority, low-income vs. high-income) and trained on local client models utilizing federated learning. A well-trained, fairness-aware “teacher” model then learns robust, group-balanced representations, and its distilled knowledge is transferred to the client models, enabling them to achieve fairer predictions without compromising overall accuracy. Numerical experiments on a large-scale travel survey dataset illustrate the effectiveness of the proposed approach to preserve overall prediction accuracy and mitigate demographic-specific unfairness.

TC61

Building B Level 4 B401

Integrative Modeling for Multi-Source and Multi-Level Data

Invited Session

Quality, Statistics and Reliability

Chair: Fenglian Pan, The University of North Carolina at Charlotte, Charlotte, NC, United States

Co-Chair: Shenghao Xia, Bowling Green State University, Bowling Green, OH, United States

Co-Chair: Jian Liu, University of Arizona, Tucson, AZ, United States

1 - Multi-Level Sensor Fusion Through Factor Graphs

Arpan Kusari, University of Michigan, Ann Arbor, MI, United States, Jiahang Li, Hojun Son

Sensor fusion presents a robust method to assimilate sensing data from diverse modalities to provide robust and accurate estimates for dynamic objects in a scene. This paper presents a comprehensive approach to sensor fusion leveraging the power of factor graphs—a highly flexible and efficient framework that represents complex estimation problems involving multiple sensors. By modeling the relationships between each sensor measurements and the states as a first level bipartite factor graph to create sensor tracks and between pairs of sensor tracks to create a fused track, our method efficiently encodes and solves estimation tasks. The proposed framework incorporates measurements from diverse sensor modalities, such as cameras, and LiDAR, and optimally fuses this information by solving for the maximum a posteriori (MAP) estimates. Through extensive experimentation across various scenarios, we demonstrate the superiority of factor graph-based sensor fusion in achieving high-accuracy state estimates with improved resilience to sensor noise and faults. This work highlights the adaptability and scalability of factor graphs as a unifying approach to sensor fusion, paving the way for future advancements in autonomous systems and intelligent perception technologies.

2 - A Weighted Dirichlet Process Mixture Model for Spatial Clustering of Child Trafficking Vulnerabilities

Wenyu Gao, UNC Charlotte, Charlotte, NC, United States

Child trafficking remains a critical social challenge characterized by complex spatial and contextual variations. This study employs a weighted Dirichlet process mixture (WDPM) model to investigate trafficking vulnerability patterns across 138 Sierra Leonean enumeration areas (EAs). Incorporating areal information from the EAs as random effects, a generalized linear mixed model (GLMM) is developed to examine the relationship between trafficking prevalence and potential risk factors. By applying the nonparametric Bayesian clustering approach to the random effects, this research innovatively identifies latent risk factors and spatial dependencies. The WDPM model is able to integrate socioeconomic and cultural factors into its weight function, enhancing clustering precision and model estimation accuracy. Unlike traditional clustering techniques, the WDPM approach dynamically determines cluster configurations without requiring predefined cluster numbers. Clustering results are rigorously validated through community-level hypothesis testing, ensuring robustness and reliability. This novel framework provides a sophisticated analytical method for understanding trafficking risks in resource-constrained settings. By uncovering latent vulnerability patterns, the study offers actionable insights for designing targeted, context-specific intervention strategies. The findings highlight a data-driven approach with significant potential to inform policy-making and optimize resource allocation to combat child trafficking effectively.

3 - Registration-Free Monitoring of Unstructured Point Cloud Data via Intrinsic Geometrical Properties

Mariafrancesca Patalano, University of Padua, Padua, Italy, Giovanna Capizzi, Kamran Paynabar

Recent advances in 3D printing technologies have enabled the production of lightweight and topologically optimized components, now widely adopted in several fields, including prosthetic design, aerospace, and automotive industries. Such innovative shapes are inspected in the form of high-dimensional unstructured point clouds of varying size. In this context, preliminary alignment of the nominal and real point clouds (i.e., registration) can be avoided by monitoring the intrinsic properties of the underlying manifold. Nevertheless, existing approaches rely on surface reconstruction and decimation algorithms that can introduce artifacts and non-manifold components, potentially affecting monitoring outcomes.

Our aim is to provide a flexible solution for inspecting arbitrarily complex shapes while avoiding surface reconstruction. To this end, we operate directly on point clouds and propose a registration-free online monitoring procedure. Our proposal consists of two alternative schemes that leverage the intrinsic properties of the shape, captured via geodesic distances and the Laplace operator. In both schemes, thresholding techniques are used to select intrinsic features most indicative of potential out-of-control conditions. Numerical studies suggest the effectiveness of the proposed procedure and its potential to identify different types of defects.

4 - Autonomy vs. Safety: Joint Modeling of Disengagement and Collision Events from Autonomous Driving Study

Simin Zheng, Virginia Tech, Blacksburg, VA, United States, Jared Clark, Jie Min, Yili Hong

As the popularity of artificial intelligence (AI) continues to grow, AI systems have become increasingly embedded into various aspects of daily life, leading to significant transformations in industries and changing how people live. One of the typical applications of AI systems is autonomous vehicles (AVs). In AVs, the relationship between the level of autonomy and safety is an important research question to answer, which can lead to two types of recurrent events data being recorded: disengagement and collision events. This paper proposes a joint modeling approach with multivariate random effects to analyze these two types of recurrent events data. The proposed model captures the inter-correlation between the levels of autonomy and safety in AVs. We apply an expectation-maximization (EM) algorithm to obtain the maximum likelihood estimates for the functional form of the fixed effects, variance-covariance components, and the tuning parameter for the penalty term. This proposed joint modeling approach can be useful for modeling recurrent events data with multiple event types from various applications. We analyze disengagement and collision events data from the California Department of Motor Vehicles AV testing program to demonstrate its application.

Key Words: AI reliability; AI robustness; Correlated frailty; EM algorithm; Recurrent events data; Survival models.

5 - Reliability Modeling for Perception Systems in Autonomous Vehicles: a Recursive Event-Triggering Point Process Approach

Fenglian Pan, University of Arizona, Tucson, AZ, United States, Jian Liu

Ensuring the reliability of sensor-fusion-based perception systems is crucial for the safe deployment of autonomous vehicles. Such systems function through a sequence of interconnected stages, where errors in upstream stages may propagate to downstream stages and trigger additional errors. The cross-stage error propagation conceptually exists and makes errors in different stages, not independent, posing model challenges, estimation challenges, and data challenges for reliability modeling. The existing methods cannot be applied to address all these challenges. Thus, this paper presents a recursive event-triggering point process to explicitly consider the error propagation based on the simulated data. The data are simulated from a proposed error injection framework, which can generate various errors from a sequence of interconnected stages in a perception system. The latent and probabilistic error propagation information is incorporated into a modified expectation-maximization (EM) algorithm for parameter estimation. The numerical and physics-based simulation case studies demonstrate the prediction accuracy and interpretability of the proposed modeling methodology.

TC62

Building B Level 4 B402

Machine Learning and Artificial Intelligence for Advanced Manufacturing

Invited Session

Quality, Statistics and Reliability

Chair: Xiaoyang Song, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Wenbo Sun, University of Michigan Transportation Research Institute, Ann Arbor, United States

1 - Quality-Ensured in-Situ Process Monitoring Using Deep Canonical Correlation Analysis

Xiaoyang Song, University of Michigan, Ann Arbor, MI, United States, Wenbo Sun, Judy Jin

In-situ process monitoring is a key challenge in additive manufacturing. Different from conventional approaches that rely solely on real-time process data, this study proposes a novel method that integrates offline quality inspection data to enable more quality-oriented process monitoring. The core objective is to uncover nonlinear dependencies between in-situ process sensing signals and offline quality outcomes. To this end, Deep Canonical Correlation Analysis (DCCA) is employed to bridge the two data modalities—process signals and quality measurement data—by extracting highly correlated latent features that capture their shared representations. These learned features are subsequently used in an online monitoring algorithm for real-time process monitoring. The proposed method is supported through a series of statistical analysis and simulation studies. Last, the proposed method is applied to a real-world metal additive manufacturing process, where optical emission spectra and CT scans are the corresponding process signals and quality data, respectively. The superior performance in this case study demonstrates its effectiveness in enhancing in-situ monitoring capabilities.

2 - Multi-Fidelity Bayesian Optimization for Process Parameter Tuning in Additive Manufacturing

Chaoran Dou, Virginia Tech, Blacksburg, VA, United States, Raghav Gnanasambandam, Mihir Darji, Prahalada Rao, Zhenyu Kong

Laser Powder Bed Fusion (LPBF) is a key technique in metal Additive Manufacturing, enabling the fabrication of geometrically complex components with high precision. However, the quality and reliability of printed parts are highly sensitive to process parameters, particularly laser power and scan speed, which govern the thermal behavior during fabrication. A critical metric in this context is the end-of-cycle temperature, as it influences residual stress accumulation, microstructure evolution, and overall part integrity. Accurately predicting this temperature typically requires high-fidelity simulations, which are computationally expensive and make parameter optimization impractical. To address this challenge, a Multi-Fidelity Bayesian Optimization (MFBO) framework is proposed for efficient and accurate process parameter tuning in LPBF. The framework employs Multi-Fidelity Gaussian Processes to integrate data from both low- and high-fidelity thermal simulations, enabling the construction of a surrogate model that captures the complex relationship between input process parameters and thermal responses while significantly reducing computational cost. The proposed approach is applied to the optimization of laser power and scan speed, with the objective of controlling the end-of-cycle temperature to reach a specified target. Results demonstrate that the MFBO framework effectively identifies optimal parameter settings using substantially fewer high-fidelity evaluations than conventional methods, without compromising predictive accuracy. This work highlights the potential of multi-fidelity optimization strategies to accelerate process parameter optimization in LPBF and offers a generalizable methodology for other simulation-driven manufacturing applications.

3 - Human-in-the-Loop Convolutional Neural Network for Gas Tungsten Arc Welding Process Control

Shenghan Guo, Arizona State University, Mesa, AZ, United States, Yue Cao, Yuming Zhang

Gas tungsten arc welding (GTAW) process control is paramount to ensure the welding quality meets the industrial standard. Due to the uncertainty and volatility associated with GTAW processes, high-fidelity modeling and system controllers have been challenging. A human expert will monitor the thermal videos collected during a GTAW process, evaluate the weld pool size and shape, and then adjust the speed parameter based on his/her knowledge and experience in GTAW to influence the weld pool and ensure it stays close to the reference. This study conceptualizes manual GTAW control as a human-in-the-loop (HiL) system and investigates HiL machine learning to develop an AI-based GTAW system controller. Conventionally, AI models must be trained on given data and then taken to real-time implementations. The training data must be cleaned and preprocessed a priori to AI model training. In this study, we propose a human-in-the-loop Deep Learning (HiL-DL) model that feeds on the human-annotated GTAW thermal images of weld pools and fine-tunes its model weights in real-time. During a GTAW process, in-situ thermal images of weld pool will be annotated with the expert-suggested speed; a pretrained ResNet50 will be integrated with a Stacked Long Short-Term Memory to learn from the annotated images and fine-tune its model weights. Upon the completion of the process, the HiL-DL model will have updated its knowledge about GTAW weld pool size, shape, and their mapping to the desirable speed. By fine-tuning such a HiL-DL in multiple GTAW processes, we expect the model to automatically recommend speed adjustment.

4 - Learning-Based Optimization for Complex Physical Systems

Yinan Wang, Rensselaer Polytechnic Institute, Troy, NY, United States

Learning-based optimization (i.e., reinforcement learning) is known for its strength in generalizability and scalability. Generalizability indicates the knowledge learned in optimizing the training data can be directly applied to optimize the unseen testing scenarios without retraining. Scalability represents the capability of handling giant design space. In this talk, we will discuss the recent research work on developing learning-based optimization for complex physical systems. We select two cases as examples. One is the fixture layout design for reducing shape deformation of large-scale sheet parts in the assembly process. The other is optimizing the control parameters in the Brookhaven Alternating Gradient Synchrotron (AGS) to improve intensity and emittance preservation after the RF bunch merge.

5 - FedScope-Kd: Knowledge Distillation-Enhanced Federated Learning via Shared Composition and Personalized Exploration for Heat Emission Prediction in Additive Manufacturing

Rong Lei, Rutgers University, Piscataway, NJ, United States, Yuebin Guo, Weihong Guo

Additive manufacturing (AM) enables the production of customized and complex metal parts. However, maintaining consistent product quality for AM processes remains challenging, which makes predicting heat emissions become essential. Small- and medium-sized enterprises (SMEs) acquire limited and heterogeneous datasets due to unique printing configurations, making traditional centralized predictive model training paradigms impractical. Federated learning (FL) offers a solution, it enables collaborative learning across decentralized datasets while keeping data local and private. However, its performance may suffer under heterogeneous client data. This creates a need to address data heterogeneity while maintaining both global model performance and individual client personalization. We propose FedScope-KD, a novel heterogeneity-aware FL framework designed to balance the scope of global and personalized learning. Client heterogeneity level is quantified prior to the FL training, to improve the model weight aggregation during training. This is achieved through component decomposition and knowledge distillation. The framework decomposes client data into process-invariant representations, which capture the fundamental nature of the laser powder bed fusion (LPBF) process, and process-variant representations, which reflect local, client-specific configuration settings. This decomposition ensures that global models benefit from collaborative learning on the homogeneous data representations, while personalized predictions remain tailored to client-specific heterogeneous data representations. The variant components exhibit greater distances between clients, reflecting their inherent heterogeneity, while the invariant components are more closely aligned across clients. We demonstrate that FedScope-KD effectively addresses the challenges of data heterogeneity, providing a robust solution of predictive modeling for SMEs.

TC63

Building B Level 4 B403

AI and Deep Learning for Supply Chains

Contributed Session

Contributed

Chair: Amir Kordijazi, University of Southern Maine, Portland, ME, United States

1 - Software Supply Chain: Investments in Cybersecurity by Firms in the Presence of Learning Cost

Siqi Li, School of Management, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Abhijeet Ghoshal, Chao Fang

The advent of AI facilitated the creation of several AI-powered tools by foundation firms, such as GitHub Copilot and Amazon Q Developer. However, these tools expose users to cybersecurity vulnerabilities passed on by the tool through the generated content due to poisoned data used during training by the foundation firm. Subsequently, the cyber criminals attack the user firm via the vulnerable code. In summary, the supply chain of AI tools passes on vulnerabilities to users via the AI tool, requiring efforts from both the foundation firm and the user to fight cybercrime. The users incur another type of cost while using AI tools, the AI tool learning cost, which is high initially and decreases with continued use. Knowing that the users face the cost of effort for fighting cyber threats and a learning cost, how much effort should the foundation firm expend on cybersecurity? Further, since it knows that the user's learning cost will eventually decrease, will it spend more or less effort than the user on cybersecurity? What happens to cybersecurity efforts when the user's learning cost decreases faster, enabling it to increase investment in cybersecurity, potentially? Are the efforts of the foundation firm and user substitutes? We answer such pressing questions in this paper using a game-theoretic model. We find that the efforts of the foundation firm and the user may be complementary under certain conditions when the users have a higher learning rate. Also, the foundation firm may invest more than the user under certain conditions.

2 - AI-Based Inventory Allocation with Fairness Constraints

Sirine Taleb, American University of Beirut, Beirut, Lebanon, Karmel Shehadeh

We present an AI-based inventory ordering and allocation system that integrates machine learning with an optimization model incorporating group-based fairness constraints. The system is designed to promote equitable decisions while providing explainable recommendations. We evaluate its performance using both synthetic and real-world demand data. Key performance metrics include cost efficiency, fairness violations, and model explainability under varying levels of demand uncertainty.

3 - Smart Retail Solutions Enabled by Generative AI and Computer Vision for Enhancing Customer Experience Management

Jongsawas Chongwatpol, NIDA Business School, National Institute of Development Administration, Bangkok, Thailand

Artificial Intelligence presents a significant opportunity to transform the retail industry. How can AI solutions be effectively designed and implemented to support and manage the in-store customer journey? This study presents a design journey for AI implementation aimed at understanding and improving the in-store customer shopping experience by integrating both Generative AI and Computer Vision AI. The design thinking process is adopted to ensure that challenges and potential failure factors in AI solutions are addressed during the proof-of-concept stage focusing on idea generation and testing.

This study presents several use cases demonstrating how Generative AI can be used to integrate the customer journey and point of view to gain insights into shopping behavior and preferences. Market analysts and store managers can then redesign the store layout by developing discrete-event simulation prototypes. The goal is to transform improvement ideas into experiments and, consequently, evaluate them through what-if scenarios.

The second objective of this study is to implement the “new store layout” prototype in a real-world setting by physically rearranging the store into different zones. Each zone is equipped with CCTV and video cameras, enabling the store to utilize data captured through Computer Vision AI to enhance information visibility. This includes tracking customer behavior at each touchpoint throughout their shopping journey—from the moment they enter the store until they make a purchase or leave. The goal is to leverage and integrate the wealth of information provided by facial recognition technologies into marketing strategies, enabling more effective personalized offers and follow-up campaigns.

4 - Lightweight Object Detection for on-Device Machine Vision

MINJONG SHIN, Kyung Hee University, Yongin City, Korea, Republic of, Seohyun Choi, Jae-Yoon Jung

The adoption of deep learning-based machine vision is rapidly advancing within the manufacturing sector. With the growing demand for real-time decision support, there is an increasing need for lightweight AI models capable of performing inference on edge devices. This study investigates the real-time object detection capabilities of a low-power device—specifically, the Raspberry Pi. The experimental setup consisted of a Raspberry Pi 5 and the Raspberry Pi Camera Module 3. A quantized YOLOv11n model was deployed on the Raspberry Pi for inference tasks. The evaluation dataset comprised images for detecting damaged areas on milling heads using machine vision techniques. TensorFlow Lite and OpenVINO served as the primary deployment frameworks for model quantization. Experimental results indicated that the average inference time was 163.5 ms in a GPU environment and 283.8 ms on the Raspberry Pi, demonstrating its suitability for real-time detection. These findings confirm that real-time vision processing is achievable on low-power platforms without significant loss in accuracy, supporting the feasibility of deploying smart vision systems under edge computing constraints.

5 - Agentic AI for Self-Refining Demand Forecasting in CPG Supply Chains

Sriram Sambasivam, Estee Lauder, EDISON, NJ, United States, Suneeth Sreedharan

The increasing complexity of Consumer Packaged Goods (CPG) supply chains—driven by product proliferation, regional heterogeneity, and omnichannel fulfillment—necessitates forecasting systems that are adaptive, modular, and capable of learning. This work presents an early-stage implementation of Agentic Artificial Intelligence (AI) within a statistical demand planning framework, wherein autonomous, goal-directed agents are employed to enhance system responsiveness and refinement.

Developed on a Databricks-native platform with a foundation in open-source libraries and composable components, the forecasting engine supports multiple planning horizons (execution, operating, strategic) and product scopes (e.g., Basic and New Product Launches). It leverages a diverse ensemble of models spanning statistical methods, tree-based algorithms, and neural networks. The introduction of agentic behaviors—such as automated forecast diagnostics, context-aware model selection, and anomaly detection—enables a progressive shift from static model execution to feedback-driven adaptation.

Rather than pursuing full autonomy, these agents emulate elements of human planning cognition: monitoring forecast performance, proposing targeted interventions, and iteratively improving based on historical signal patterns. Their decisions are explicitly linked to business-aligned performance metrics, including weighted forecast accuracy (WFA), plan attainment, completeness, and forecast bias.

This work outlines the modular architecture, key implementation considerations, and guiding design principles for embedding agentic intelligence within enterprise-scale forecasting systems. It represents a pragmatic, scalable approach toward intelligent automation in supply chain planning—one that balances technical rigor with operational viability.

6 - Explainable Graph Neural Networks for Anomaly Detection and Disruption Propagation in Supply Chain Networks

Samridhi Vats, Amazon, Bellevue, WA, United States, Swapnali Karvekar

Supply chains are inherently graph-structured systems, yet most anomaly detection models treat them as flat or time-series data—ignoring the cascading and relational nature of disruptions. This work introduces a novel application of **Graph Neural Networks (GNNs)** to detect and localize anomalies in complex, multi-tiered supply chain networks.

Unlike traditional models that rely on tabular inputs or sequential patterns, our framework captures the **relational dependencies** between entities—such as suppliers, distribution hubs, and logistics routes—by learning from both **node-level behaviors** and **edge-level interactions**. We evaluate our approach on publicly available datasets, including **SupplyGraph** and a synthetic anomaly-injected version of the **H&M Fashion** dataset. These graphs represent real-world supply networks where disruptions, delays, or abnormal flows can propagate in non-obvious ways.

To enhance interpretability and address industry concerns around deep learning black-box models, we incorporate **GNNExplainer-based modules** to generate node- and edge-level attributions for flagged anomalies. Our framework aims to bridge the gap between advanced AI techniques and practical, risk-aware supply chain operations.

This research contributes to the growing intersection of operations research and machine learning by proposing a scalable, explainable, and graph-native anomaly detection pipeline. It lays the groundwork for a new class of AI tools capable of **early warning, root-cause tracing, and proactive mitigation** in supply chain risk management.

Keywords:

Graph Neural Networks, Supply Chain Resilience, Anomaly Detection, Explainable AI, Network Analytics, Machine Learning

7 - A Performance Comparison of State-of-the-Art Object Detection Algorithms and Their Sensitivity to Image Characteristics and Performance-Measuring Toolkits

Mehrsa Mashhadi, University of South Florida, Tampa, FL, United States, Nicolas Bustos, Susana Lai-Yuen, Sudeep Sarkar, Tapas Das

The task of object detection in computer vision is highly important for many applications including surveillance systems, autonomous driving, medical imaging, etc. Recently, multimodal object detection has gained significant attention due to its ability to maintain high accuracy under challenging environmental conditions, including adverse weather and low light scenarios by leveraging complementary information from multiple imaging modalities such as RGB and thermal images. However, evaluating the performance of these algorithms using varying custom datasets and metric tools has made consistent and fair comparisons challenging. In this research, we conduct a systematic evaluation of four state-of-the-art object detection algorithms—Detection Transformers, Faster R-CNN, YOLOv7, and YOLOv9—on the publicly available Aligned FLIR dataset, which provides aligned RGB and thermal image pairs. Each algorithm was trained and tested on RGB, thermal, and RGB-Thermal fused images, which were generated using TarDAL, a feature-level fusion algorithm. We investigated each model's detection performance and sensitivity to crucial factors including image modality, illumination conditions (day vs night), different object sizes, and image resolution variations. In addition, we analyze the impact of using different metric calculation frameworks on reported mean average precision (mAP). Our results indicate that YOLOv7 outperforms other detection algorithms in all the scenarios. Among different image modalities, thermal images yield the highest detection performance. Also, it highlights how the choice of evaluation toolkit can change the relative ranking of detectors. Our research proposes standardized evaluation guidelines to promote fair and reliable comparisons on future studies in the field of multimodal object detection.

8 - Accelerated Material Innovation: An AI-Driven Approach Using Bayesian Optimization

Amir Kordijazi, University of Southern Maine, Portland, ME, United States

This study examines how Bayesian Optimization (BO) can enhance materials design and manufacturing by guiding experimental processes to achieve specific optimal material properties. In materials design, achieving high-quality and consistent outputs involves navigating a vast parameter space, and BO offers a targeted approach to do this efficiently. By prioritizing experiments that yield the most valuable data, BO refines manufacturing settings iteratively, significantly reducing the time and resources needed to optimize results across diverse materials.

In addition to improving experiment selection with Bayesian Optimization, automation streamlines the testing process, helping researchers and manufacturers reduce hands-on tasks and enhance workflow efficiency. This study presents an adaptable, AI-driven framework that powers a self-driving lab, facilitating rapid and precise tuning of manufacturing parameters. This approach enhances productivity and accelerates the development of optimized material properties for specialized applications. It highlights the role of AI and automation in fostering faster innovation in manufacturing and advancing R&D in material design by simplifying the discovery of optimal configurations for high-performance, application-specific materials.

TC64

Building B Level 4 B404

Equity, Institutions, and Social Impact in Economic Systems

Contributed Session

Contributed

Chair: Salma Shaik, University of Pittsburgh, 300 Campus Dr, Bradford, PA, 16701, United States

1 - The Impact of Accommodation Choices on The Feeling of Loneliness: a Surrogate Method for Measuring Association and Moderation for Discrete Choice Variables

Zhaohu Fan, Scheller College of Business, Georgia Institute of Technology, Atlanta, GA, United States, Shaobo Li, Lvy Liu, Philip S. Morrison, Dungang Liu

We propose a new residual-based method for regression models analyzing discrete choice outcomes. This work extends existing methodologies for discrete data analysis, building on frameworks introduced by Liu and Zhang (2018) and Liu et al. (2021). More recently, Li et al. (2024) developed a framework for quantifying and visualizing partial associations between mixed variable types. While their work has advanced our understanding of such relationships, the COVID-19 pandemic has intensified interest in examining how accommodation choices relate to loneliness, particularly in the context of remote learning among college students. Our study focuses on the association between accommodation types (e.g., student hall, home with parents, or renting) and the feeling of loneliness, an ordinal measure of emotional well-being. A key statistical challenge arises when one of the variables—such as accommodation—is a multinomial variable with no inherent order. To address this, we introduce a $[0,1]$ -scaled measure of association, where 0 indicates no association and 1 indicates perfect association. Our approach assigns orders to the unordered categories and computes the association measure T_i for each. The maximum of these values serves as the overall association measure. This framework also supports covariate adjustment, allowing us to estimate partial associations. Preliminary findings show that the method can effectively quantify the influence of a discrete choice variable on another outcome, offering a clear and interpretable summary measure. Our approach fills a methodological gap in analyzing associations involving unordered categorical variables and enhances interpretability in complex social science applications.

2 - Beyond Gdp: COVID-19'S Effects on Macroeconomic Efficiency and Productivity Dynamics in OECD Countries

Umit Saglam, East Tennessee State University, Johnson City, TN, United States

This study analyzes macroeconomic productivity dynamics across 37 OECD countries from 2018 to 2024, focusing on the impact of the COVID-19 pandemic and government response measures. Using a Slack-Based Measure Data Envelopment Analysis (SBM-DEA) and the Malmquist Productivity Index (MPI), we assess changes in total factor productivity (TFP) across pre-, during-, and post-pandemic periods. Our framework incorporates both desirable (GDP) and undesirable outputs (inflation, unemployment, housing price inflation, and interest rate distortions), offering a multidimensional view of macroeconomic efficiency. Results show broad but uneven productivity gains, with technological progress proving more resilient than efficiency during the pandemic. Post-COVID recovery trajectories diverged, reflecting differences in structural adaptability and innovation capacity. Regression analysis reveals that stringent lockdowns in 2020 were associated with lower productivity in 2023–2024, while more adaptive policies in 2021 supported long-term technological gains. These findings highlight the importance of aligning crisis response with forward-looking economic strategies and demonstrate the value of DEA-based methods for evaluating macroeconomic performance beyond GDP.

3 - Explaining Health Disparities via Accessibility: a Data-Driven Machine Learning Framework Using Shap

Jiewen Luo, university of Massachusetts Amherst, Amherst, MA, United States, Eric Gonzales

Understanding the impact of transportation access is pivotal for improving public health outcomes. However, the variety of transportation modes, community amenities, and the interactions between accessibility mode and opportunities often obfuscate the comprehensive understanding of the contributors to health disparities. This study clarifies the concept of accessibility and introduce a novel approach to measurement walking, biking, and public transportation accessibility by incorporating both mode

adoption probability for residents and built environment characteristics. We propose a framework employing three interpretable model (Elastic Net, XGboost, Geographical Random Forest (GRF)), enables detailed exploration of the relationship between accessibility and public health outcomes (obesity, diabetes, and heart disease) at both global and local scales. A case study has been conducted in Massachusetts. Key factors are interpreted using Shapley Additive Explanations (SHAP), which also facilitate the synthesis of multiple forms of accessibility into a single aggregated measure reflecting their contribution to health disparities. The proposed framework aims to provide more comprehensive insights into how improving access can enhance public health, identifies the types of accessibility most strongly associated with specific the health outcome, and explores how these relationships vary by demographic factors such as age, income, and car ownership. Results indicate that GRF outperforms other models when spatial autocorrelation (global Moran's I) is high and variable importance varies geographically. Obesity has a better fit on linear model, while hypertension and heart disease have non-linear relationship with accessibility. Findings suggest that targeted improvements in critical access through infrastructure planning can efficiently improve public health outcome.

4 - Exploring the Causal Inter-Relationships Among Arrests, Officers, and Crime Rates in the United States

Salma Shaik, University of Pittsburgh at Bradford, Bradford, PA, United States, Lakshika Kuruppuarachchi

The unexpected crime drop phenomenon in the early 1990s garnered increased interest to investigate the effects of various factors on crime but detailed studies analyzing the causal relationships between arrests, officers, and crime rates at different geographical levels are lacking. This study uses a comprehensive dataset from 1974-2020 to address this gap. Panel ARDL and Granger Causality techniques are utilized to estimate the short-run and long-run coefficients and to identify causal relationships by accounting for cross-section dependency, non-stationarity, slope heterogeneity, and co-integration. Analysis is carried out for violent crimes and property crimes separately for very large, large, and medium cities which are categorized based on the FBI guidelines. Results demonstrate the presence of both short-term and long-term relationships between arrests and crime across all agencies. Granger causality tests reveal that property arrests exhibit bidirectional causality with both property and violent crimes. Also, violent crimes drive violent arrests up especially in large and medium agencies while officers affect the outcome of crimes for all agencies. These results have significant policy implications since they show that though arrests and officers have deterrence effects on crime, they are not substantial. It is critical to ensure that everyone has access to good education, employment opportunities, and affordable housing so that the incentive to commit crime is reduced. Increased efforts should be made to reduce the rate of recidivism among repeat offenders. This study serves as a valuable resource for policymakers to understand the complex crime dynamics and to devise effective crime prevention strategies accordingly.

TC65

Building B Level 4 B405

AI Reasoning and Behavioral Simulation

Invited Session

Artificial Intelligence

Chair: Marios Papachristou, Cornell University, Ithaca, NY, 14850, United States

Co-Chair: Caitlin Cunningham, University of Washington, Seattle, United States

1 - Agreeing to Disagree: Human-AI Collaboration in Ethical Decision-Making

Caitlin Cunningham, University of Washington, Seattle, WA, United States, Rebekah Baik, Leonard Boussieux, Leela Nageswaran

Though AI systems increasingly provide decision-support in subjective scenarios, their influence in ethically ambiguous decisions remains poorly understood. We conducted a lab experiment with participants evaluating ethical dilemmas grounded in real, everyday situations to examine how large language model (LLM) judgments, varying in decisiveness and justification, shape user agreement, confidence, and perceived helpfulness. We find evidence that users weigh AI input significantly more than their personal input, especially when reasoning is provided. Our findings also reveal that decisive AI recommendations increase user alignment and confidence, while justifications from a reasoning model enhance perceived helpfulness. These effects vary by the controversy level of the anecdote: minimal advice suffices in clearer dilemmas, while the added justification improves persuasion in more complex ones. Our findings offer actionable guidance for deploying LLMs in subjective human decision-making contexts where trust, deliberation, and user agency are at stake.

2 - Conversation Dynamics in Human-AI Collaborative Teams: Study of Group Decision-Making in Mixed-Agent Environments

Jiehui Luo, University of Notre Dame, Notre Dame, IN, United States, Margaret Traeger, Ahmed Abbasi

As generative AI systems enter collaborative environments, understanding their impact on team dynamics in complex decision-making tasks becomes crucial. While research has focused on one-to-one human-AI interactions, significant gaps exist in understanding how these dynamics extend to group settings. Our research examines human-AI collaboration in small groups (triads) performing complex evaluation tasks like essay grading. We vary generative AI interaction modalities and agency to study their influence on conversation patterns and outcomes. Using multi-method conversation analysis, we quantify how generative AI integration affects social dynamics.

Preliminary findings reveal that when generative AI systems engage in discussions containing negative emotional content, they produce larger evaluation gaps compared to human graders. This sensitivity to conversational affect creates generative AI judgment patterns that influence group evaluation processes and outcomes. Additionally, we observed that the cognitive processing level of generative AI contributions evolves across discussion rounds - initially displaying higher analytical complexity before gradually shifting toward simpler consensus-building language in later exchanges. Our data comparing final scores after pure generative AI setting discussions shows varying levels of alignment with human evaluators across different models. Our ongoing research examines various human involvement configurations, testing how expertise distribution, role clarity, and team composition influence evaluation processes and deference to generative AI judgments. These insights aim to provide guidance for designing generative AI systems that can be effectively deployed in educational assessment contexts while preserving the diversity and depth of human judgment.

3 - Too Much of a Good Thing? the Hidden Cost of Widespread AI-Generated Persuasion

Yan Leng, University of Texas at Austin, Austin, TX, United States

The rapid adoption of large language models (LLMs) has transformed digital communication, enabling users across commerce, fundraising, political campaigns, and other sectors to generate persuasive text at scale. While this AI-driven productivity boost appears beneficial, it introduces a novel platform-level externality: As users independently employ similar LLM prompts to optimize their content, messages increasingly resemble one another, leading to audience fatigue and diminished collective effectiveness. We term this phenomenon stylistic congestion. In a formal congestion model, users adopt overlapping stylistic "bits," whose persuasive value diminishes as usage increases. The resulting strategic environment forms a potential game in which uncoordinated LLM use generates substantial welfare losses that grow with user scale and adoption rates. This digital tragedy of the commons arises from the success and scale of generative AI. Our findings suggest that sustaining collective persuasive effectiveness and maximizing welfare will require platforms to shift from passively enabling AI delegation to actively governing it.

4 - Measuring Strategization in Recommendation: Users Adapt Their Behavior to Shape Future Content

Hannah Li, Columbia, New York, NY, United States

Most modern recommendation algorithms are data-driven: they generate personalized recommendations by observing users' past behaviors. A common assumption in recommendation is that how a user interacts with a piece of content is a reflection of the content, but not of the algorithm that generated it. Although this assumption is convenient, it fails to capture user strategization: that users may attempt to shape their future recommendations by adapting their behavior to the recommendation algorithm. We test for user strategization by conducting a lab experiment and survey. To capture strategization, we adopt a model in which strategic users select their engagement behavior based not only on the content, but also on how their behavior affects downstream recommendations. Using a custom music player that we built, we study how users respond to different information about their recommendation algorithm as well as to different incentives about how their actions affect downstream outcomes. We find strong evidence of strategization across outcome metric. For example, participants who are told that the algorithm mainly pays attention to "likes"/"dislikes" use those functions 1.9x more than participants told that the algorithm mainly pays attention to dwell time. A close analysis of participant behavior rules out experimenter demand as the main driver of these trends. Further, in our post-experiment survey, nearly half of participants self-report strategizing "in-the-wild," with

some stating that they ignore content they actually like to avoid over-recommendation of that content in the future. Our findings suggest that user strategization is common and that platforms cannot ignore the effect of their algorithms on user behavior.

5 - Llm Routing with New Models Through Active Learning

Alkiviadis Mertzios, Massachusetts Institute of Technology, Cambridge, MA, United States, Wei Sun, Georgia Perakis

LLM routing is an emerging area focused on selecting the best large language model for each query based on predicted cost and output quality. Prior work assumes a fixed pool of models and trains routing policies accordingly. We extend this setting to a more realistic scenario where new language models, such as domain-specific finetuned models, are introduced after the initial training phase. Since evaluating every query on the new model is expensive, we propose an active learning approach to selectively acquire the most informative queries. This enables efficient integration of new models with minimal evaluation cost.

TC66

Building B Level 4 B406

Behavioral and Market Analytics

Contributed Session

Contributed

Chair: Hanyu Zhang, Emory University, Decatur, GA, United States

1 - Cultural Dynamics in Online Consumer Behavior: Exploring Cross-Country Variances in Review Influence

Eunjung Lee, Lewis University, Romeoville, IL, United States

This study explores how cultural differences shape online consumer behaviors, specifically investigating how national cultural characteristics influence consumers' expressions in online reviews. Guided by Hofstede's Cultural Dimensions Theory, the research compares consumer reviews from two culturally distinct countries—the United States and South Korea—examining cultural dimensions such as Individualism vs. Collectivism, Power Distance, and Uncertainty Avoidance. A mixed-methods approach employing text mining, data mining, and topic modeling is utilized to analyze review data collected from popular platforms like Netflix and Agoda.com. This methodology allows the identification of culturally-specific patterns, such as individualistic consumers (U.S.) focusing on personal experiences and direct critiques, whereas collectivist consumers (Korea) emphasize communal experiences and group harmony. Furthermore, the study investigates whether high power-distance cultures show greater deference and politeness, compared to low power-distance cultures which display more openness to direct critique. To ensure accurate and culturally unbiased analysis, a two-way translation process is adopted, translating English reviews to Korean and vice versa. Ultimately, this research seeks to validate and extend Hofstede's cultural dimensions theory in digital consumer contexts. The findings will provide valuable insights for multinational companies, enabling them to tailor their marketing strategies, customer engagement, and review management systems effectively across different cultural markets. This contributes significantly to academic literature on cultural psychology and digital consumer analytics, offering actionable recommendations for businesses aiming to better resonate with culturally diverse global audiences.

2 - Between Originality and Popularity: How to Successfully Create Digital Content

Kang Lee, Fairfield University, Fairfield, CT, United States, Minjee Sun, Matthew Osborne

As digital content consumption becomes more routine, a range of content creators, from novices to professionals, are increasingly engaged in creating digital content. Despite the proliferation of digital content, few pieces of work achieve success and financial rewards. For a content creator to achieve a successful outcome, the content should follow a popular trend to capture the attention of potential customers, and also for algorithms to expose it to customer searches. At the same time, the content must have some originality: in other words, it must have some distinct features that ensure it can maintain customer attention in the long term. This research explores how content creators balance popularity and originality in their content creation. We utilize natural language processing (NLP) and text mining to identify popular and trending topics, then employ graph theory to construct network graphs for these topics, assessing the novelty and originality of content aligned with specific trends. Essential outcome measures, such as view counts and subscriptions, guide the determination of the optimal blend of popularity and originality. Our preliminary findings reveal that the emphasis on novelty varies significantly depending on the content's topic. Consumers exhibit preferences for stereotypical content in specific topics, lean toward balanced content in others, and seek highly distinctive content within each topic. Next, we plan to extend our findings to diverse digital content to ensure generalizability. Our research can assist content creators in developing more engaging and effective content and support digital platform managers in identifying and promoting promising content creators.

3 - Promotion Optimization for Durable Goods

Ronak Ramvishal Tiwari, WHU - Otto Beisheim School of Management, Vallendar, Germany, Arnd Huchzermeier

The profitability of promotions is often a concern for retailers. Management scholars have episodically studied the problem of determining profit-optimal promotional prices for CPG and fast-fashion sectors, much less so - empirically. We utilize data from an online bike retailer and study promotional pricing in the less-discussed durable goods context.

4 - Review Sentiment Analysis - Based Recommendation System

Soyeon Kim, Ewha Womans University, Seoul, Korea, Republic of, Dongil Kim

Recommendation systems are essential tools for delivering personalized experiences that enhance customer satisfaction and drive conversion rates. Existing conventional approaches that rely primarily on explicit rating data often fail to capture the nuanced emotional context behind user preferences. In contrast, user-generated reviews provide rich contextual information – sentiment and specific product-related preferences – that extend beyond numerical scores. This research employs natural language processing techniques to extract both emotional signals and product-related preferences with textual reviews. By incorporating these insights into the recommendation process, the research aims to develop a more context-aware that better reflects user intent, ultimately enabling more personalized and meaningful recommendations.

5 - Taste or Terminate? Understanding Business Competitiveness Through Consumer Cohort Reviews

Jie Li, University of Massachusetts Lowell, Malden, MA, United States, Asil Oztekin

For businesses with a significant offline presence, such as restaurants and hotels, understanding the role of competitive positioning in business survival is essential. However, assessing competitiveness remains challenging in today's dynamic market environment. This study examines key determinants of competitiveness—business feature proximity, geographical density, and consumer reviews—and their impact on restaurant survival. Through leveraging publicly available data from social media platforms which cover over 9,000 restaurants from 2005 to 2021, we apply survival analysis and panel regression framework to incorporate time-lagged consumer opinions and competitive attributes. The findings indicate that restaurants with less differentiation from their competitors face a higher likelihood of closure. Over time, consumer reviews have a long-lasting effect and continue to influence restaurant survival, contributing to accumulated reputational effects. Additionally, new customer reviews and sentiment have a stronger impact on competitive positioning than those from returning customers, highlighting the importance of first-time consumer experiences. These insights help inform competitive strategies and support industry practitioners in decision-making, emphasizing the role of consumer perception management and differentiation in maintaining business sustainability.

6 - As Transactions GO Social: Predicting User Engagement Through Language and Network Evolution with Dynamic Gnn

Hanyu Zhang, Emory University, Atlanta, GA, United States, Pantelis Loupos, David Schweidel

Venmo, the leading peer-to-peer payment platform, has profoundly transformed social and financial interactions by allowing users to observe friends' transaction activities within a social feed. By leveraging the diffusion of Venmo, we aim to examine the dyadic interactions among users, and unpack the impacts of network structure, content

(i.e., texts and emoji associated with transactions), and historical behaviors on consumer engagement in a dynamic network setting. We propose a theory-driven dynamic graph neural network (GNN) architecture that integrates both structural and communication data to effectively predict user engagement. The model leverages textual and emoji analyses to detect shifts in communication patterns and relational dynamics. It also adapts to changes in users' hidden statuses and enriches neighborhood message passing with data from social feeds. Evaluated within a live-update setting for dynamic link prediction, our GNN demonstrates that incorporating textual data significantly enhances performance, improving predictive accuracy by 7.72% compared to non-textual baseline models. This research provides a robust framework for understanding and predicting the complex behaviors of users on digital transaction platforms, highlighting the importance of integrating social feed dynamics into engagement strategies in evolving network environments.

TC67

Building B Level 4 B407

Advanced Machine Learning for Complex Physical Systems

Invited Session

Data Mining

Chair: Mucun Sun, Idaho National Laboratory, Idaho Falls, 83401, United States

Co-Chair: Yinan Wang, Rensselaer Polytechnic Institute, 110 8th St, Troy, 12180-3522, United States

1 - Advanced Machine Learning for 3D Polymer Structure Recognition with Explainable AI Analysis Advanced Machine Learning for 3D Polymer Structure Recognition with Explainable AI Analysis

Zhenyu Kong, Virginia Tech, Blacksburg, VA, United States, Amirul Islam Saimon, Soumil Joshi, Jihoon Chung, Sanket Deshmukh

Bottle Brush Polymers (BBPs), with their densely grafted side chains, show promise in applications like drug delivery, smart coatings, and advanced filtration. Their performance depends heavily on their 3D shape, typically captured through Molecular Dynamics (MD) simulations as point cloud data. Machine learning models, particularly 3D convolutional neural networks (CNNs), have been used to classify these shapes, but require computationally expensive voxelization of point clouds—both during training and inference. In response, this paper proposes a PointNet-based model that recognizes temperature-sensitive BBP shapes directly from raw point cloud data without the need for voxelization. Additionally, to make the shape predictions explainable to domain experts, we apply critical-bead method and gradient-based techniques that highlight the specific polymer beads most responsible for model's recognition decision. By combining model performance and explainability, we provide a practical tool that enables the rational design of advanced polymer materials.

2 - Causality-Informed Machine Learning in Manufacturing

Xubo Yue, Northeastern University, Jamaica Plain, MA, United States

The manufacturing world runs on data—sensors, logs, and real-time feeds—but are we really learning from it, or just drowning in noise? Machine learning (ML) has become the industry's favorite hammer, smashing through problems like predictive maintenance and quality control. But too often, these models chase correlations instead of *true causes*, leading to brittle solutions that fail when the real world throws a wrench in the gears.

Enter *causality*. Forget "good predictions"—what if we could uncover *why* a machine fails, *how* a process drifts, or *which* adjustments actually matter? A handful of bold researchers are already doing this, blending causal inference with ML to build models that don't just predict—they *understand*. Think Bayesian networks exposing hidden bottlenecks, or counterfactual simulations testing process changes before they're made. This isn't just academic fluff; it's the difference between a model that works on paper and one that survives the factory floor.

In this talk, we'll tear apart the hype, spotlight real-world wins (like causal ML slashing defect rates in automotive welding), and ask the hard question: *If causality is this powerful, why isn't every manufacturer using it?*

3 - PgcF-Net: a Prototype-Guided Causal Fusion Network for Interpretable Multivariate TIME-Series Anomaly Detection

Jahoon Jeong, Yonsei University, Seoul, Korea, Republic of, Hyunsoo Yoon

Time series anomaly detection often suffers from limited interpretability, as deep models typically operate as black boxes. We present PGCF-Net, a novel multivariate time-series anomaly detection (MTSAD) model that balances high detection performance with interpretability. The model extracts four types of anomaly evidence—predictive deviation, causal graph changes, contextual deviation, and spike magnitude—via a prototype-guided and causally-aware evidence extractor. These are adaptively fused by a learned evidence fusion network. Trained in two unsupervised stages, PGCF-Net achieves state-of-the-art results on five public benchmarks while providing rich, actionable explanations to support root cause analysis.

4 - Learning to Collect: a Two-Stage Data Collection Framework for Data-Efficient Personalization

Shuai Huang, University of Washington, Seattle, WA, United States, Feng Lin

Collaborative learning (CL) has been developed as a generic framework to develop personalized prediction models. Existing scope of CL focuses on learning from data collected from a heterogeneous collection of individuals, and has not considered how to improve the quality of the collected data. In this work, built on the foundation of CL, we propose Learning to Collect (LtC), a two-stage framework for efficiently collecting data from individual users. We theoretically and empirically demonstrate our design can induce personalized models to achieve competitive performance with less data compared to a full data design.

TC68

Building B Level 4 B408

Innovation in Market Design and Revenue Management

Invited Session

Revenue Management

Chair: Ozan Candogan, University of Chicago, Chicago, IL, United States

Co-Chair: Tonghun Lee, University of California, Berkeley, Berkeley, CA, United States

Co-Chair: Yiding Feng, Hong Kong University of Science and Technology, Hong Kong, Hong Kong

1 - Time-Based Contracts in Spatial Networks

Tonghun Lee, University of California, Berkeley, Berkeley, CA, United States, Ozan Candogan, Yiding Feng

Digital freight brokerages connect shippers and carriers through online platforms, but inefficiencies arise due to empty backhaul miles when return shipments are unavailable. These inefficiencies persist due to drivers' preference for cyclic routes that start and end at the same location, yet they often face difficulty securing return loads, potentially reducing their long-term participation on the platform. To address this, we propose a time-based mechanism design in which the platform assigns cyclic routes to drivers, who are compensated based on their on-duty time. We first characterize the driver equilibrium through a convex program formulation of a generalized population potential game and show its existence and essential uniqueness. Our approach to time-based mechanism design centers on a bi-level optimization problem, where the platform seeks to maximize its profits under driver equilibrium. To address the large-scale non-convex setting, we propose a first-order method that leverages a reformulation of the equilibrium characterization as a regularized optimization problem. Using this framework, we conduct numerical studies based on U.S. freight data to

assess the impact of time-based contracts on spatial supply-demand imbalances, platform profit, and driver surplus. We also examine how network expansion decisions influence platform profit.

2 - Information Design In Research Funding: Threshold Mechanisms and Performance Guarantees

Mengyang Lin, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Anran Li, Wenqi Lian, Zhe Liu

Research funding agencies must allocate limited funding opportunities among researchers who strategically disclose information about their projects' true quality, resulting in a novel mechanism design challenge. We model this scenario as an information design problem with multiple senders (the researchers) and a single receiver (the funding agency). Each project's value comprises a public component (θ), reflecting the researcher's track record, and a private component (η), representing the current project's quality. Beginning with a single-researcher setting, we characterize the researcher's optimal disclosure strategy under a threshold-based selection mechanism, which is shown to be optimal and achieves the full-information welfare.

Extending to the multi-researcher setting introduces competitive dynamics that render the threshold-based mechanism suboptimal in general. Moreover, identifying a truly optimal mechanism is computationally infeasible, and we demonstrate that funding projects purely based on the highest posterior values can fail to admit an equilibrium. Consequently, we refine the threshold-based policy to accommodate a capacity constraint of k maximum funding slots. Our policy guarantees a performance ratio of $\frac{1 - \sqrt{\frac{1}{k+3}}}{2}$ relative to the full-information optimum, inducing a simple equilibrium that efficiently allocates funding. The results provide actionable insights for designing research funding mechanisms under strategic information disclosure and limited resources.

3 - Learning and Dynamic Equilibria in Misspecified Duopoly Pricing

Feiyu Han, University of Chicago, Chicago, IL, United States, Ozan Candogan

We study a misspecified multinomial logit (MNL) duopoly Bertrand pricing model in which each firm treats itself as a monopolist. Firms use a stochastic gradient-based online learning algorithm with myopic monopoly pricing, following the approach in Goyal et al. (2023). While prior work shows convergence to the unique Nash equilibrium (NE) when price sensitivity in the utility is fixed, we relax the fixed price sensitivity assumption and allow it to vary over positive values. This flexibility introduces richer dynamics: the learning process does not necessarily converge to the NE. Instead, we introduce the notion of a dynamic equilibrium (DE): a pricing profile at which firms' expected decisions no longer change over time under the learning dynamics. We analyze the associated continuous-time ODE and show that multiple DEs may arise. Although the NE qualifies as a DE, it is not distinguished—it lies in the interior of the revenue set, defined as the set of all feasible revenue pairs. This implies the NE does not generally maximize revenue for either firm. Finally, we show that the NE is not a stable DE. These results highlight a gap between static equilibrium concepts and long-run outcomes under misspecified learning, and show that even small changes in utility specification can lead to qualitatively different dynamics.

4 - Learning a Stackelberg Leader's Incentives from Optimal Commitments

Yurong Chen, INRIA, Paris, France

Stackelberg equilibria, as functions of the players' payoffs, can inversely reveal information about the players' incentives. In this paper, we study to what extent one can learn about the leader's incentives by actively querying the leader's optimal commitments against strategically designed followers. We show that, by using polynomially many queries and operations, one can learn a payoff function that is strategically equivalent to the leader's, in the sense that: 1) it preserves the leader's preference over almost all strategy profiles; and 2) it preserves the set of all possible (strong) Stackelberg equilibria the leader may engage in, considering all possible follower types. As an application, we show that the information acquired by our algorithm is sufficient for a follower to induce the best possible Stackelberg equilibrium by imitating a different follower type. To the best of our knowledge, we are the first to demonstrate that this is possible without knowing the leader's payoffs beforehand.

5 - A System-Level Analysis of Conference Peer Review

Yichi Zhang, Rutgers University, New Brunswick, NJ, United States

Conference peer review involves three constituencies with different objectives: authors seek fast acceptance at prestigious venues, conferences want to present a program with many high-quality and few low-quality papers, and reviewers want to avoid excessive workloads. These objectives are far from aligned, primarily because paper evaluation is inherently noisy. Over the years, conferences have adopted various policies to manage these tradeoffs, such as adjusting acceptance thresholds, varying the number of reviews per submission, requiring prior reviews for resubmissions, and others. In this work, we investigate the effectiveness of these policies and the reasons behind their success or failure.

We model the conference-author interactions as a Stackelberg game: a conference commits to an acceptance policy; the authors best-respond by (re)submitting or not (re)submitting to the conference in each round of review, the alternative being a "sure accept" (such as a lightly refereed venue). Our analysis yields several insights. First, the conference should typically set a higher acceptance threshold than the actual desired quality --- creating a "resubmission gap". Second, the reviewing load is heavily driven by resubmissions of borderline papers - therefore, a judicious choice of acceptance threshold may lead to fewer reviews while incurring an acceptable loss in conference quality. Third, conference prestige, reviewer inaccuracy, and author patience increase the resubmission gap, and thus increase the review load for a fixed level of conference quality. For robustness, we further consider different models of paper quality and compare our theoretical results to simulations based on plausible parameters estimated from real data.

TC69

Building B Level 4 B409

AI and Operations

Invited Session

Applied Probability Society

Chair: Hongseok Namkoong, Columbia University, New York, NY, United States

1 - Simulating and Experimenting with Social Media Mobilization Using LLM Agents

Mohsen Bayati, Stanford University, Stanford, CA, United States, Sadegh Shirani

Social networks have transformed the ways in which political mobilization messages are disseminated, raising new questions about how peer influence operates at scale. We develop an agent-based simulation framework that integrates real demographic data, Twitter network topology, and heterogeneous LLM agents to examine the effect of mobilization messages on voter turnout. This provides a controlled, reproducible environment for testing counterfactual designs in political mobilization research.

2 - AI and Operations: A View to the Future

Srikanth Jagabathula, NYU Stern School of Business, New York, NY, United States

This is a high level talk that covers how traditional research within operations interacts with transformative technologies, such as large language models, generative AI, and agentic systems. The focus will be on taking a view to the future and identifying interesting research opportunities.

3 - Discrete Optimization for Large Language Model Compression

Kayhan Behdin, LinkedIn, Mountain View, CA, United States

Large Language Models (LLMs) have demonstrated remarkable performance across a wide range of industrial applications, from search and recommendations to generative tasks. Although scaling laws indicate that larger models generally yield better generalization and performance, their substantial computational requirements often render them impractical for many real-world scenarios at scale. In this work, we present methods and insights for compressing LLMs into Small Language Models

(SLMs) that maintain the generalization performance of the LLM, while requiring less inference compute resources. Particularly, we focus on a key technique known as structured pruning, discussing algorithms rooted in discrete optimization. Additionally, we share the lessons we have learned from applying model compression techniques based on discrete optimization in a large-scale industry use case.

TPC

Building A Level 1 Exhibit Hall A3

Tuesday Poster Competition Session

Poster Session

Contributed

1 - Feedforward Neural Network (FNN) for Real-Time Packaging Strength Prediction

Supanart Sukrason, The Pennsylvania States University , Malvern, PA, United States, Animesh Agrawal, Rayat Noor

This study presents a Feedforward Neural Network (FNN)-based classification model for real-time strength prediction in corrugated box packaging, aiming to eliminate production latency caused by traditional quality assurance (QA) procedures. Conventional mechanical testing requires operators to halt production, extract sample boxes, and send them to a QA lab for compression testing—a process that takes approximately 10 minutes per lot. This delay results in machine idle time and throughput reduction, contributing to estimated annualized losses exceeding \$2.7 million.

In contrast, the proposed system enables inline, automated strength classification directly from live process parameters, eliminating the need for sample extraction or manual testing. The model is trained on industrial-grade data from SCG Packaging, incorporating multivariate features such as edge crush test (ECT) values, print cylinder gaps, additive formulation types, machine speed, box geometry, flute type, and etc. The model result is defined as binary outcomes (“Pass” or “Not Pass”) for supervised classification.

Five FNN architectures were developed and evaluated using performance metrics including accuracy, precision, recall, F1-score, and AUC-ROC. The optimal model (Model 0) achieved an AUC of 0.998 and accuracy of 98.6%, with minimal false positives/negatives.

Moreover, a real-time data pipeline integrates machine telemetry and raw material data via APIs and a middleware layer, supporting low-latency inference. This development enables continuous model predictive quality assurance, reducing downtime and improving operational efficiency.

This work highlights the role of deep learning and industrial AI in operations research, offering a scalable framework for intelligent, real-time quality control in packaging manufacturing.

2 - Causal Analysis for Congressional Trade - ML, Vector Db and AI Agents at Work

Hari Hara Prabhu Narayanan, dotData, Austin, TX, United States

We present a novel causal inference framework that integrates machine learning, hybrid vector search, and AI-driven investigative agents to evaluate the market impact of congressional stock trades. Our method identifies comparable control stocks using multi-modal embeddings (Vector DB), estimates counterfactual price trajectories through Bayesian structural time series (BSTS) modeling, and leverages LLM-based agents to autonomously investigate external policy or financial events that may validate causal relationships.

Analyzing over 10,000 congressional disclosures from 2022–2024, we find that 34% of trade clusters exhibit statistically significant post-trade deviations, often corroborated by AI-augmented external investigations. Our scalable and interpretable framework offers a new pathway for uncovering how political activity may causally influence financial markets, with implications for transparency, policy analysis, and investment strategy.

3 - Moveod: a Data Pipeline for Synthesizing Origin-Destination Demand Data Using U.S. Census Information

Rishav Sen, Vanderbilt University, Nashville, TN, United States, Samitha Samaranayake, Ayan Mukhopadhyay, Aron Laszka, abhishek dubey

Detailed, time-indexed origin–destination demand matrices provide the empirical foundation for transportation network optimization, vehicle-routing algorithm design, and congestion management studies in operations research. Unfortunately, publicly available OD data for the United States are usually coarse, mode-specific, or subject to disclosure limits that preclude detailed analysis. We present MoveOD, an open-source, fully automated pipeline that synthesizes time-stamped, building-level commuter flows for any United States county while exactly reproducing official Census marginals.

MoveOD statistically fuses four data streams: (i) Longitudinal Employer–Household Dynamics OD flows (LODES), (ii) American Community Survey (ACS) departure-time histograms, (iii) ACS travel-time distributions, and (iv) building-level footprints from OpenStreetMap and Microsoft Buildings Footprint. An integer-linear program calibrates candidate trips so that three constraints hold simultaneously: origin–destination totals, origin by departure-block counts, and origin-specific travel-time histograms. The resulting synthetic commuters are spatially and temporally realistic yet privacy-preserving.

To illustrate MoveOD’s utility, we generate about one million synthetic trips for six southeastern U.S. counties, but the same pipeline applies to any county nationwide. These building-level, time-stamped data can feed directly into downstream studies such as benchmarking vehicle routing heuristics, evaluating congestion pricing, or stress testing multimodal network designs, analyses that are impractical with conventional zonal demand.

To facilitate adoption, we will release the complete, documented codebase with a web-based interactive interface through which practitioners select a county, set temporal parameters, and download calibrated OD tables together with the accompanying metadata and shapefiles. MoveOD enables rapid, reproducible experimentation for researchers, planners, and analysts across transportation, logistics, and public policy.

4 - Returns Consolidation and Outbound Freight Optimization: a Two - Stage MIP Model for Reverse Logistics

Sheetal Sheetal, University of Iowa, Iowa City, IA, United States

The convenience of e-commerce has led to a surge in online returns, introducing significant logistical and operational challenges for reverse logistics networks. Offline stores generate high volumes of returns that must be consolidated, sorted, and shipped back to merchants through multi-tier supply chains. Delays in consolidation, inefficient hub routing, and suboptimal freight mode choices can significantly increase operational costs. Designing coordinated shipment policies under stochastic return volumes and service-level constraints is thus a critical operational challenge.

We develop a mixed-integer programming model to optimize these operations for third-party logistics service providers managing e-commerce returns. The model captures consolidation at return drop-off stores, shipment routing to hubs, freight mode selection (parcel, LTL, and FTL) for dispatches to merchant warehouses, and information coordination between return points and hubs.

Our experiments simulate 100 days of operations across multiple stores, hubs, and merchants, with Poisson-distributed demand at each store. We test multiple scenarios involving box size flexibility, store-to-hub assignments, and hybrid shipping policies that combine quantity-based triggers with minimum service frequency guarantees.

Results show that hybrid policies —where shipments are triggered based on accumulated volume but must occur at least once a week—can yield cost savings of up to 30% over rigid time-based schedules while maintaining service levels. In contrast, allowing multiple box sizes for shipping from drop-off locations to hubs provides minimal additional savings. These findings highlight the importance of synchronization between returns consolidation and freight dispatch decisions, and offer practical insights for third-party providers designing scalable and cost-efficient reverse logistics systems.

5 - Strategic Placement of Drone Platforms in Out-of-Hospital-Cardiac-Arrest: A Case Study for Jefferson County, Louisville, Ky

Marziehsadat Rezaei, University of Louisville, Louisville, KY, United States, Monica Gentili

Each year, over 350,000 out-of-hospital cardiac arrests (OHCA) occur in the U.S., with survival rates as low as 8–10% due to delayed ambulance arrivals and limited use of publicly available Automated External Defibrillators (AEDs). Unmanned Aerial Vehicles (UAVs), or drones, present a transformative solution for rapid AED delivery to OHCA scenes, overcoming barriers like AED accessibility and bystander hesitation. Using Louisville's EMS data, this study identifies priority areas for drone delivery and potential launch sites in Jefferson County, Louisville, KY. These inputs inform an optimization model to determine the minimum number of launch platforms needed to cover a specified percentage of intervention calls within a desired service time. We present results of the optimization model under different scenarios, by varying the priority areas, the number and location of the platforms, and the drone flight parameters.

6 - A Roadmap to Sustainability for University Chemical Labs

Xinyue Zhang, Boston University, Boston, MA, United States, Xindi Liao, Chenxi Feng

This research developed a roadmap to help Boston University labs achieve My Green Lab Certification. Issues found in audit processes identified as duplicate purchases and data gaps. Recommendations include greener chemicals, optimized procurement processes, and equipment alternatives. A fluid dynamic model and energy efficiency estimates project sustainability improvements.

7 - Development of a Virtual Reality-Based Quadcopter Simulator for Immersive Training and Research in Rescue and Emergency Missions

Emre Eraslan, University of Illinois Urbana Champaign, Urbana, IL, United States, Avinash Gupta

Unmanned Aerial Vehicles (UAVs) are increasingly used in high-risk, time-sensitive scenarios such as search and rescue or disaster response. While autonomous UAVs have advanced, effective human oversight remains essential in complex, dynamic environments. Existing training tools often lack realism and fail to place the operator in immersive, decision-making roles. To address these limitations, this project introduces a digital twin-inspired virtual reality (VR) simulator for swarm-based UAV operations.

The system integrates MATLAB for control modeling and Unity for immersive visualization. Completed components include a nonlinear UAV dynamics model, closed loop reference model-based adaptive controllers for leader-follower formations, and autonomous obstacle avoidance via simulated sensors. A joystick interface enables real-time teleoperation of the leader UAV. A key technical achievement is the establishment of two-way data communication between MATLAB and Unity, allowing dynamic interaction and real-time feedback.

The ongoing work involves developing a VR environment with the XR Interaction Toolkit, terrain features, and embedded rescue mission scenarios. Human factors principles will inform interface design to reduce cognitive load and improve usability. The system will undergo pilot testing and structured usability evaluations.

This simulator offers a novel training and research platform that combines high-fidelity control, swarm dynamics, and immersive user interaction. Beyond operator training, it supports the study of human-system performance under mission-critical conditions. Long-term, this work lays the foundation for real-world UAV swarm deployment through VR-based interfaces. Potential applications include emergency response, digital mapping, and autonomous inspection, where coordinated autonomy and human oversight must seamlessly coexist.

8 - Dynamic Robust Integer Optimization with Cost Uncertainty

Pengyu Zhang, University of Michigan, Ann Arbor, MI, United States, Ruiwei Jiang, Guodong Lyu, Stanley Lim

We study a dynamic robust integer optimization model with cost uncertainties, characterized by stagewise Wasserstein balls. The model seeks to find a dynamic policy for discrete decisions, with an objective of minimizing the expected total cost under the worst-case distribution. We recast the model as a classical robust dynamic integer program by proving a minimax theorem and explicitly characterizing the worst-case distribution. In addition, we design a dual dynamic programming algorithm to solve this model, alongside lower and upper bounds that converge to the optimal value almost surely. Furthermore, we investigate value of multistage modeling (VMS) by benchmarking our model with popular alternatives, including a two-stage model and a myopic policy. In this analysis, we discover that VMS is particularly significant when the uncertainty has either a high variance or a high heterogeneity over time. The results from extensive numerical experiments in a real-world last-mile delivery and a route planning problem validate our findings.

9 - The Hidden Cost of Politeness: Evaluating The Environmental and Operational Costs of AI Prompt Design

Victoria Kehinde, Georgia State University, Atlanta, GA, United States, Mahira Begum, Lauretta Fagbuyi, Juliet Dokor, Nesreen El-Rayes, Wendy Roth

As artificial intelligence systems, particularly large language models (LLMs), become integral to daily digital interactions, seemingly benign user behaviors, such as adding polite phrases like “please” and “thank you,” may carry unexpected consequences for sustainability. This study investigates the relationship between prompt structure and energy consumption, examining how the inclusion of polite or verbose language affects processing time, compute intensity, and operational costs at scale. Using benchmark tests and industry energy use data, the research quantifies the marginal energy and cost increases associated with longer, more nuanced prompts and explores their cumulative environmental impact across millions of daily interactions.

The study also evaluates how prompt optimization strategies and user education might mitigate computational waste without sacrificing conversational quality. This study provides new insight into the intersection of human behavior, digital efficiency, and sustainability. It raises critical questions about the hidden costs of user interaction design in AI systems and suggests pathways for more energy-conscious AI use. The findings have implications for data center resource planning, responsible AI design, and digital sustainability practices, offering a novel lens through which to assess the societal and operational impacts of language in computational systems.

10 - Co-Optimizing Spacecraft Routing and Depot Location-Allocation in Cislunar Space

Mustafa Tanbir Kuhel, Texas State University, San Marcos, TX, United States, Tongdan Jin

The space industry is expected to grow at an exponential rate, reaching \$2.7 trillion by 2045, necessitating innovative supply chain frameworks to enable long-term deep space explorations. This research tackles the unique issues of space logistics by co-optimizing spacecraft routing, orbital depot placement, and rocket staging for human residence on lunar surface. Our Space Depot Location-Allocation (SDLA) model uses mixed-integer programming to reduce overall launch mass while maintaining depot capacity, flight route, propellant technologies, and orbital mechanics. The model evaluates single- and two-stage spacecraft design by applying the Tsiolkovsky rocket equation to compute fuel and structural mass for orbital depots and delivery routes from Earth to Moon.

Numerical experiments for Earth-Moon network of 6 candidate orbital depots, and 12 lunar bases reveal that two-stage rockets reduce total launch mass by 4,949 kg, equivalent to \$10.11M cost savings compared to single-stage systems, achieving a 5.05% fuel reduction. Sensitivity analysis identifies that raising specific impulse from 450s to 600s reduces mission mass by 23.3%, while lowering structural coefficients results in mass savings of up to 5.5%, emphasizing the importance of advanced propulsion and structural design. Siting depot at gravitationally stable nodes, such as Lagrange orbits, improves logistical efficiency in the cislunar network.

Future additions will include in situ resource utilization, additive manufacturing, and interplanetary flight time constraints in SDLA model. The framework offers valuable insights for reducing costs and improving logistics sustainability for lunar explorations in the next decade, aligning with global efforts to establish permanent human

habitats on the Moon.

11 - A Milp Approach for Multi-Period Order Allocation Under All-Unit Quantity Discounts and Blending Ratios

Fuhad Ahmed Opu, University of Texas at San Antonio, San Antonio, TX, United States, Moddassir Khan Nayeem, Hamid Najafzad, Yara Altarawneh, Omar Abbaas

This study addresses optimal order allocation under all-unit quantity discounts for manufacturing a single product requiring multiple ingredients with fixed blending ratios. Suppliers' quantity-based discounts introduce non-linearities, complicating multi-period cost minimization involving procurement, inventory holding, and fixed ordering costs under deterministic demand. We propose a MILP model based on a finite set of Critical Order Quantities (COQs) that preserves optimality and eliminates the need for nonlinear discount modeling. Mathematical propositions validate that ordering at these COQs efficiently minimizes total costs. This framework supports strategic procurement decisions, effectively balancing immediate cost savings and long-term inventory impacts.

12 - City Observation and Learning: a Map Tile Based Spatial Data Fusion Pipeline

Wing Au, Fujitsu Research of America, Santa Clara, CA, United States, Xiaojian Zhang

We present the pipeline to transform heterogeneous geospatial data into a common format and be combined or 'fused' for downstream tasks. Leveraging a convention that is used by Web Map technology, the format can be easily interpreted and controlled by laymen decision makers due to its uniform, intuitive representation. The mechanism addresses three main obstacles making geospatial data difficult to use in a widespread manner: multi-spectrum, generalizability and transferability. 2D/3D spatial images, 2D geometric shapes, 1D temporal data from various sources such as demographics, land-use, point-of-interest, weather, etc. are seamlessly assimilated as uniform features for the data pipeline. Observations in various social economics use cases such as traffic, public safety, city operation and environmental impact are analyzed within a generalized framework. Using global or national data sources as a basis, analyses and learnings performed on one locality can be transferred to a similar locality, where data may not be as abundance. A GUI-based application complements the data pipeline enhances usability by providing visual comparison of the various features and observations to produce the most statistically useful combination for downstream simulation or Machine Learning and Artificial Intelligence usage. Our goal is to target business and municipal operations, where a common framework can be applied anywhere, any problem with any data.

13 - Finite Time Analysis of the Natural Policy Gradient in Finite-Horizon Markov Decision Processes

Asha Barua, Virginia Polytechnic Institute & State University, Blacksburg, VA, United States, Sajad Khodadadian

The natural policy gradient is a well-known algorithm for finding the optimal policy in a Markov Decision Process (MDP). Variants of this algorithm, such as TRPO and PPO, have demonstrated remarkable performance in real-world applications. This success has sparked extensive research on the finite time analysis of these algorithms. In this work, we present the first finite time analysis of the natural policy gradient for finite-horizon MDPs where H is the length of the horizon and k is the iteration, establishing an $\mathcal{O}(H^2/k)$ rate of convergence.

14 - Growth of Income Funds and Death of Volatility : a Study in Inelastic Option Supply and Market Impact

Abdurrahman Kurucak, University of Texas at Dallas, Richardson, TX, United States, Taeyoung Park

Over the past decade, income funds employing the call overwriting strategy have emerged as one of the fastest-growing segments in the mutual fund industry. This study explores the consequential impact of their trading on the underlying assets and options. By using detailed option position data of all mutual funds from N-PORT disclosure, we find that the aggregate selling of call options by these income funds reduces subsequent monthly option implied and realized volatility. Its significance stems from income funds, which repetitively overwrite calls on a regular basis rather than engaging in informed selling for volatility timing. This reduces the endogeneity concern and bolsters the importance of the inelastic supply. In the context of demand-based asset pricing, this is a critical implication because this finding proves that asset prices can vary by inelastic supply of call options. In addition, in the context of investor protection, this study reveals that this activity can detrimentally impact potential profits, as lower implied volatility diminishes the option premium and consequently reduces the cash flow stream for investors of income funds.

15 - Next-Gen Grid Security Platform

Alaa Selim, University of Connecticut, Storrs, CT, United States

The rapid digitalisation of power distribution grids exposes millions of edge-connected devices to sophisticated cyber-physical threats. I introduce the **Next-Gen Grid Security Platform**, a closed-loop framework that first identifies the grid's most damaging attack surfaces, then detects ongoing intrusions, and finally responds with optimal corrective actions. A Bayesian optimisation engine emulates a non-cooperative adversary, learning worst-case ramp attacks that maximise voltage violations and loss of critical load on the IEEE-123 feeder, outperforming genetic algorithms and CMA-ES in both convergence speed and impact. For defence, two complementary controllers are trained: a deep-reinforcement-learning volt/VAr agent that dispatches DER reactive power in real time, and a Bayesian game-theoretic optimiser that co-ordinates topology reconfiguration with DER set-points under uncertainty. Together they restore up to ninety-five percent of lost load and return voltages to within ANSI limits in three 20-ms control cycles. Detection is handled by a hybrid large-language-model text validator that screens command streams and a lightweight ResNet classifier that flags anomalous SCADA waveforms with ninety-eight percent accuracy and only twenty milliseconds of latency, satisfying NIST CSF response timelines. All layers are validated through high-fidelity OpenDSS simulations and a hardware-in-the-loop microgrid at UConn's Center for Clean Energy Engineering. The talk opens with a concise survey of emerging cyber threats in active distribution networks and demonstrates how an optimisation-centric, machine-learning-driven framework embedded at the protective-relay layer can neutralise those threats, provide step-by-step deployment guidance for utilities, and chart future directions in robust optimisation, safe reinforcement learning, and trustworthy AI for critical-infrastructure security.

16 - An Analysis of AI-Based Methods for Visual Sensory Trigger Detection in Sensory Processing Disorder

Yara Altarawneh, The University of Texas at San Antonio - Department of Mechanical, Aerospace, and Industrial Engineering, San Antonio, TX, United States, Omar Abbaas, Leslie Neely, Moddassir Khan Nayeem, Fuhad Ahmed Opu, Hamid Najafzad

Sensory Processing Disorder (SPD) is a condition that causes difficulty in capturing, modulating, interpreting, or organizing sensory stimuli. Despite its impact on daily function, SPD continues to receive less attention in both research and practice compared to physical disabilities. Current assessments rely heavily on subjective tools, including checklists and observational reports. In this project, AI-based methods were reviewed and compared for their effectiveness in detecting visual sensory triggers, with the goal of improving sensory accessibility. A systematic review was conducted on 68 studies published between 2014 and 2025. A structured filtering process was followed, and models were evaluated based on four criteria: accuracy, processing speed, model complexity, and their ability to detect brightness, color, and contrast. The most effective techniques which outperformed traditional models in both detection capability and adaptability were identified and used in an example included in the project showing how AI can be used to process visual input by detecting visual triggers. Most of the reviewed work focused on classification and diagnosis, while visual trigger detection received limited attention. A noticeable overlap was found between studies on autism, neurodiversity, and accessibility, but direct attention to visual overstimulation in SPD remained limited. The results confirm that certain models are more suitable than others for detecting sensory triggers, and that detection-focused frameworks are still lacking. Future work should be directed toward building AI systems that detect and respond to visual overstimulation to support inclusive and adaptive environments.

17 - Who is the Better Bettor: a High School Student Using AI or a Professional Gambler?

Samuel Goldfeder, Highland Park High School, Highland Park, IL, United States, James DiLellio

This paper explores the performance of an AI-powered sports betting model compared to traditional market expectations and expert picks. Using machine learning—specifically, mass web scraping and gradient-based optimization—the model gathers a wide range of data to inform its decisions. Inputs include team efficiency metrics, historical betting lines, injury reports, public sentiment, and sharp money indicators, all used to generate probabilistic forecasts for NCAA men's basketball games.

The model learns from past results, gradually identifying which factors—like efficiency ratings, line movement, or injuries—have the greatest impact on outcomes. Over a five-week period, we tested the model on 68 games and compared its results to sportsbook closing lines and expert consensus.

The model consistently outperformed both sets of expectations, particularly in underdog and small-market matchups, where oddsmakers tend to be less accurate. It posted a 60.3% win rate, on against-the-spread (ATS) and total points scored (O/U) — well above the 50% market baseline and 57.5% expert average win rate—and returned a net gain of 12.91 units on total bets of 78.3 units. Importantly, bet size recommendations were dynamic, based on the model's confidence in each pick.

While its edge diminished closer to game time, aligning with market efficiency theory, the model proved especially effective in identifying early value. These findings suggest AI models can offer a measurable advantage in low-liquidity markets when powered by real-time data. This work adds to the growing field of algorithmic wagering and highlights the potential of adaptive, data-driven strategies in practical sports betting.

Tuesday Poster

Building A Level 1 Exhibit Hall A3

Tuesday Poster Session

Poster Session

1 - Dual Decomposition for Coordinated Optimization of Coupled Traffic Power Systems with EV Charging and Queue Dynamics

Hamidreza Attari, University of California, Davis, DAVIS, CA, United States

The integration of electric vehicles (EVs) into urban systems has created a tightly coupled relationship between transportation and power infrastructures. We propose a dual decomposition-based optimization framework that models the coordination between these two subsystems, incorporating EV routing, queue-aware charging behavior, and distribution-level power constraints.

The model includes multiple types of shared information between subsystems, specifically, EV charging demand and charging station queue length, as coupling variables. Charging decisions are influenced by Locational Marginal Prices (LMPs), which are computed as dual variables of the nodal power balance constraints in the power model and passed iteratively to the traffic subsystem as real-time price signals. EV battery dynamics are captured through SOC update constraints, ensuring physical feasibility over time.

To evaluate the coordination mechanism, we implement a decentralized solution approach using dual decomposition, enabling iterative subsystem optimization with synchronized dual updates. Crucially, we assess the relative importance of each type of shared information by analyzing the magnitude and entropy of the associated dual variables. This allows us to quantify which information contributes most to system-wide coordination and cost reduction.

Numerical experiments on a multi-node coupled traffic–power testbed demonstrate the benefits of information-aware coordination, showing improved load balancing, reduced total cost, and mitigation of charging congestion. The proposed framework provides a scalable and interpretable foundation for operating interdependent infrastructure systems under high EV penetration.

2 - Hybrid Model for High-Precision Fake Account Detection on Social Media

Kimia Keshanian, University of Tampa, Tampa, FL, United States

Fake social media accounts on platforms like X and Instagram pose serious threats, spreading misinformation, malware, and scams. This study presents a multi-level model to detect such accounts with high precision, integrating deep learning and human feedback. Using the Cresci-2017 dataset of 6,825 accounts, our approach processes numerical metrics (e.g., follower counts), bio text, profile images, and tweets through distinct layers. The first layer extracts 7 numerical features and applies TF-IDF to bios, yielding 1,000 text features. The second layer employs a convolutional neural network (CNN) to analyze 2,024 profile images, generating 64 features. The third layer uses a long short-term memory (LSTM) model to evaluate 4,285 tweet sets, producing a probability score. Combining these 1,072 features, a Random Forest classifier achieves 99.77% accuracy in 5-fold cross-validation, outperforming a numerical-only baseline (99%). Feature importance analysis shows numerical metrics contribute 55%, images 29%, tweets 12%, and bios 4%. One uncertain account (59% fake probability) is flagged for human review, enhancing practical deployment. Unlike traditional machine learning methods, our multi-level framework leverages deep learning to address prior limitations, offering a scalable, robust solution. The model and features are archived for future research. This study highlights the effectiveness of layered, human-augmented modeling in identifying fake accounts, advancing trust and security in social media platforms.

3 - Trilevel Network Interdiction Game for the Minimum Cost Flow Problem with Input Dependence

Adam Rumpf, Florida Polytechnic University, Lakeland, FL, United States, Hemanshu Kaul

We consider a game theoretic network flow model for planning optimal fortifications to a collection of interdependent civil infrastructure networks against a targeted attack or worst-case natural disaster. The model consists of a three-stage sequential game played on a generalized minimum-cost flow network with binary input dependencies, wherein certain arcs within the network become unusable if there is shortfall in delivery of flow to certain nodes (for example, modeling links in a telecommunications network that require input from nodes in the electric power network in order to be usable). The game consists of a defender selecting a subset of arcs to protect, then an attacker selecting a subset of undefended arcs to destroy, and finally the defender optimizing the flow in the damaged network, with the overall goal of choosing the initial defensive strategy that minimizes the final flow cost over all possible attacks. The presence of binary interdependencies complicates the solution process due to the introduction of integer constraints and the possibility of cascading failures. We present exact and computationally efficient approximate solution algorithms for the resulting trilevel mixed integer-linear program, including algorithms that utilize a linear relaxation of the binary interdependencies.

4 - Large Language Models in Academia: Boosting Productivity but Reinforcing Inequality

YoungJin Kwon, Carlson School of Management, University of Minnesota, Minneapolis, MN, United States, Agnes Yang

Large language models (LLMs) have garnered significant attention for their potential to enhance knowledge worker productivity. In this study, we provide the first large-scale empirical evaluations of LLMs' impact on academic research productivity. Leveraging a comprehensive dataset of 4,582 computer science scholars across 194 top U.S. universities and analyzing 251,124 research papers published between 2018 and 2024, we find that the introduction of LLMs is associated with about 8% increase in publication output—a gap that persists across alternative measures, including the first-author publications and top-tier conference papers. Our regression discontinuity in time (RDiT) analysis further reveals that LLMs not only shifted the average publication level but also accelerated the growth rate of productivity, rising to 3.2% in 2023 and 12.8% in 2024. Notably, junior scholars realize stronger gains than their senior counterparts, with the productivity benefit diminishing by roughly 1% for each additional year of experience. Recognizing that LLMs' benefits may not be uniformly distributed, we also investigate their impact on non-native English-speaking (NNEs) researchers, who have historically faced disadvantages in academic writing (Liao et al., 2024). Difference-in-differences and generalized synthetic control analyses indicate that, following LLM adoption, native English-speaking (NES) researchers produced more papers than their NNE counterparts. Overall, our findings indicate that while LLMs significantly boost scholarly productivity, they also exhibit dual effects, lowering barriers for junior scholars while potentially reinforcing linguistic inequities.

5 - Domain-Sensitive Opinion Leadership in Pandemic Communication: Identifying Influencers on Social Media

Aram Bahrini, University of Illinois at Urbana Champaign, Champaign, IL, United States, Behnam Malmir, Hossein Amini

Effective public health communication during pandemics is often hindered by misinformation and declining trust in official sources. This study proposes the Pandemic Opinion Leader Analysis (POLA) framework to identify domain-sensitive opinion leaders (OLs) in online social networks, enabling health authorities to strategically disseminate accurate and timely information. Leveraging both implicit and explicit data from X (formerly Twitter), the framework integrates social network analysis and topic modeling to uncover influential individual and organizational actors across thematic subdomains—such as vaccination, health policy, and community action.

The study identifies key leadership traits (trust, expertise, uniqueness, innovativeness, reputation) and quantifies their relevance via targeted surveys. Results reveal that OL influence is largely unidirectional, highlighting hierarchical dynamics in online discourse. The findings emphasize the importance of domain-specific OLs—individuals who exert targeted impact in niche pandemic topics—offering a nuanced approach to influencer engagement.

The POLA framework offers a scalable, data-driven tool for optimizing communication efforts in health crises and other public emergencies, contributing both theoretically and practically to information systems, digital leadership, and crisis management.

6 - Operational Design for Modular Transit Shuttle Systems During Pandemics

Xinke Cui, Drexel university, Philadelphia , PA, United States

Transit agencies need to control infection risks while maintaining service quality and lowering operational costs during global pandemics. To fulfil this need, various adaptation strategies have been proposed for transit operations under pandemic circumstances, but their reliance on fixed-capacity vehicles limits their effectiveness. In this study, we examine the use of Modular Autonomous Vehicles (MAVs) that enable variable capacity operations via autonomous docking and undocking maneuvers. To facilitate the design of transit operations with MAVs, we develop an optimization framework that simultaneously determines dispatch headways, vehicle formations, and onboard flow control levels. The consideration of both variable capacity operations and infection risks in this optimization framework poses tremendous computational challenges. To address these challenges, we reformulate the problem as a Mixed Integer Linear Programming Model for exact solution and propose a Continuum Approximation (CA) algorithm for near-optimal solution by decomposing the problem into a set of analytically and separately solvable subproblems. Numerical experiments using real traffic data from the Beijing subway system show that: (1) MILP was able to solve exact optimal solutions for small instances within 1831.11 sec; (2) the CA algorithm achieved near-optimal solutions for large instances with negligible error (0.52%) within only 5.9 sec. Results also demonstrate the potential of MAVs in reducing total system cost, despite the increase in the infection cost. Overall, the proposed model and algorithm offer transit agencies a practical decision-support tool for designing cost-effective and infection-aware operations under future pandemic conditions.

7 - Addressing Highly Imbalanced Clinical Data in Machine Learning for Cognitive Impairment Prediction Among Chronic Pain Patients

Do Kieu Trang Thoi, Virginia Tech, Blacksburg, VA, United States, Huaiyang Zhong

Chronic pain significantly impacts millions, often co-occurring with debilitating cognitive impairment. Early identification of patients at risk for cognitive decline is crucial for timely intervention and improved quality of life. This study aimed to develop and evaluate a model for predicting cognitive impairment in individuals with chronic pain. The dataset comprised electronic health records from Carilion Clinic, including demographic variables, medication history, diagnosis codes. We compared the performance of multiple classifiers, employing techniques to mitigate class imbalance. The findings demonstrate the feasibility of predictive modeling for early identification of cognitive impairment in chronic pain patients and underscore the importance of addressing data imbalance. Implementing such models in clinical decision support systems may facilitate timely interventions and improve cognitive health outcomes in this vulnerable population.

8 - Optimal Decision Support Systems for Adaptive Radiation Therapy

Fatemeh Nosrat, Rice University, Houston, TX, United States, Raul Garcia, Aysenur Karagoz, M. Soheil Hemmati, S. Mohammad Hosseinian, Andrew Schaefer, Clifton David Fuller

The purpose of this study was to determine personalized optimal timing for re-planning in adaptive organ-at-risk-sparing radiation therapy under limited re-planning resources in patients with head and neck cancer (HNC). A novel Markov decision process (MDP) model was developed using HNC patients treated at the University of Texas MD Anderson Cancer Center to determine optimal timing of re-plannings based on the patient's expected toxicity, characterized by normal tissue complication probability for four toxicities: xerostomia, dysphagia, parotid gland dysfunction, and feeding tube dependency at 6 months post-treatment. Optimal re-planning strategies were obtained when the permissible number of re-plannings throughout the treatment was to 1, 2, and 3. In contrast to prior work that relies on a single re-planning allowance or predetermined time intervals using a one-size-fits-all approach, the MDP model proposed in this paper offers a personalized, resource-aware, and scalable decision-making tool to identify optimal dynamic re-planning schedules, tailored to individual patients's needs.

9 - Information Sharing Strategies for Pharmaceutical Supply Chains

Kevin Carr, Northeastern University, Boston, MA, United States, Jacqueline Griffin

Pharmaceutical supply chains (PSCs) are critical healthcare infrastructure frequently compromised by fragmented information exchange, causing inefficiencies, stockouts, and patient safety risks. We introduce a comprehensive framework to address challenges to information sharing in PSCs through four interconnected aims. First, we identify and taxonomize unique barriers to information sharing, revealing that trust, transparency, and regulatory complexity present the most significant challenges. Second, we analyze information sharing practices from adjacent industries (e.g., retail, manufacturing, oil & gas) to identify transferable solutions. Our cross-industry matrix mapping demonstrates how specific practices can address different barrier dimensions, with the most successful solutions targeting trust and confidentiality issues. Finally, we explore modeling methods to examine how a multi-dimensional strategy addressing technical, behavioral, and organizational factors simultaneously, demonstrating that designing effective information-sharing platforms for PSCs may be less intractable than anticipated.

10 - Revitalizing Institutional Journalism: A Blockchain and AI-Driven Framework to Counter Digital Polarization

Mohamed Abouzahra, California State University, Monterey Bay, Seaside, CA, United States

The decline of traditional institutional journalism in the digital era has contributed to the rise of misinformation and the fragmentation of public discourse. This paper introduces a novel system that harnesses blockchain and artificial intelligence (AI) to restore the relevance, financial sustainability, and societal impact of institutional journalism. Leveraging a design science methodology, the study explores how blockchain-enabled smart contracts and AI-driven personalization can bridge the gap between legacy media and digital audiences on social platforms.

A large-scale analysis of over 80,000 tweets related to journalism revealed a strong correlation between users' political leanings and their preferred news sources, with content shared by individuals receiving greater engagement than posts from institutional outlets. This highlights the challenges traditional journalism faces in being perceived as external to online communities and its struggle to maintain audience trust and reach.

To counter these dynamics, we developed a system that personalizes news delivery via smart contracts on the Ethereum blockchain. AI agents analyze user interests to recommend and summarize content, which can then be purchased and shared seamlessly on social media. This community-centered approach increases article visibility, generates revenue for outlets, and reduces the barriers imposed by rigid subscription models.

The system was piloted with university students and demonstrated promising results in usability, engagement, and the fostering of democratic discourse. By embedding institutional journalism within the digital community fabric, this model presents a scalable, ethical pathway for combating polarization and revitalizing the role of journalism in democratic societies.

11 - A Three Stage Feature Selection Framework for Causal Inference

Tianyu Yang, Northeastern University, Boston, MA, United States, Alam Md Noor E

Feature selection is important in causal inference to obtain unbiased estimates of causal quantities. Selecting variables that achieve unbiased and robust causal estimation is pivotal. Recent research indicates that balancing treatment-associated variables alone introduces bias, while including spurious variables increases variance. Addressing this challenge, we propose a three-stage framework that significantly enhances feature selection compared to existing methods, yielding lower bias and variance in causal estimation while maintaining oracle property. Evaluation on synthetic datasets demonstrates superior performance across various scenarios, ensuring scalability for large datasets. Finally, we apply our methodology to real-world data, evaluating a key US healthcare policy on the causal link between opioid use disorder and suicidal behavior in the context of the opioid epidemic crisis.

12 - Sustaining the Future: A Model of Long-Term Environmental Investment in a Decentralized Supply Chains

Mukul Kaushik, University of Illinois Urbana Champaign, Urbana, IL, United States

This paper presents a model of pollution prevention in a supply chain where firms face environmental risks and increasing long-term costs from pollution. The supply chain consists of suppliers with different planning horizons – some focus only on short-term profits, while others consider medium or long-term outcomes. Pollution events occur randomly and lead to private penalties as well as externalities that grow over time, capturing rising reputational damage and regulatory compliance costs. Firms can choose to form an alliance to co-develop a pollution-reducing technology, adopt the technology by paying a licensing fee, or do nothing. Adoption decisions are not verifiable, allowing for the possibility of greenwashing. A downstream retailer sets wholesale prices and contract durations, which shape the incentives for upstream firms. The model captures how firm decisions depend on planning horizons, technology costs, pricing, and the accumulation of pollution-related compliance costs. It provides a structured way to analyze when and how decentralized supply chains invest in long-term environmental sustainability, even without strict regulation or enforcement.

13 - A Fairness Framework for Resource Substitution in Large Networks

Ved Mohan, Georgia Institute of Technology, Atlanta, GA, United States, El Mehdi Er Raqabi, Pascal Van Hentenryck

A major challenge for many organizations operating large-scale networks is ensuring that the right resource is available at the right location and at the right time. This is an NP-hard problem because uneven demand patterns and the subsequent asymmetric flow of resources across the arcs often create imbalances on the nodes of these networks. To deal with these imbalances, a cost-effective approach is to perform resource substitution among multiple, potentially composite and interchangeable, resource types. This leads to the resource substitution problem, a very challenging task in both theory and practice, which aims to find the minimum number of resource substitutions from the initial resource assignment in the network so that the overall imbalance is minimal. However, resource substitution cannot be satisfactory unless the preferences of various network schedulers are taken into consideration. This paper presents a generic framework that combines operations research (OR) and machine learning (ML) to ensure a fair resource substitution within large networks. The OR component allows modeling and solving the resource substitution problem under a fairness lens. The ML component leverages the historical data to learn schedulers' preferences, intelligently explore the decision space, and enhance the OR component's effectiveness. The framework outputs a portfolio of solutions from which the schedulers can pick satisfactory solutions. We apply our framework to the large network of a leading US parcel carrier, which motivated this research. Experimental results demonstrate the potential benefits of the developed framework against existing state-of-the-art approaches to resource substitution in the literature.

14 - Eliciting An Expert's Confidence in Their Chosen Model Construction Assumptions

Jennifer Pazour, Rensselaer Polytechnic Institute, Troy, NY, United States, Patrick Brewer

All models are wrong, some are useful. One reason even high-fidelity models are wrong is that they rely on humans to make translations from the system they are modeling to mathematical or computational form. While humans typically have some beliefs about which model construction assumptions are least like reality versus the assumptions in which they have the most confidence, this information is typically discarded once the model is put into use.

Thus, in this work, we explore how to design elicitation protocols to capture an expert's confidence in their chosen model assumptions and inputs. Grounded by a survey of existing expert elicitation protocols, this work identifies a portfolio of protocols that can elicit expert judgements and their confidence in their judgements at the model assumption-building stage. Follow on work translates this highly valuable knowledge into a method to quantify a predictive model's representative ability.

15 - Mindful Interactions with AI: a ChatGPT Case Study

Lu Xiao, Arizona State University, Tempe, AZ, United States, Marina Mitiaeva

Mindful communication involves being present, attentive, kind, and non-judgmental—expressed through ethical awareness, interpersonal sensitivity, and responsiveness. As generative AI systems like ChatGPT increasingly function as communication partners, a key question emerges: Can they engage in mindful communication—and does it matter? To explore this, we analyzed 48,000 real Human–ChatGPT conversations, examining how morality, politeness, and alignment—three core dimensions of mindful communication—unfold in practice. We also conducted a qualitative classroom study to investigate how prompt tone influences learning experiences.

Our findings reveal that while ChatGPT consistently employs moral and polite language, it lacks adaptive flexibility, limiting its responsiveness to user context and relational nuance. In contrast, human users demonstrate greater emotional and lexical variability, highlighting a gap in reciprocal adaptation.

We propose a layered framework of mindful AI communication: anchored in moral grounding, shaped by politeness, and realized through alignment. Crucially, these dynamics suggest that users increasingly perceive AI not merely as a tool, but as a social presence. Designing AI, therefore, is not just a technical endeavor—it is the crafting of social identity. This work offers both conceptual insight and practical guidance for developing AI systems that are socially intelligent, ethically grounded, and capable of nurturing trust, empathy, and sustainable human–AI relationships.

16 - Quick Recovery of Feasible MINLP Solutions from MILP Relaxations for Water Network Throughput Maximization

Sai Krishna Kanth Hari, Los Alamos National Laboratory, Los Alamos, NM, United States, Russell Bent

Determining the maximum demand a water distribution network can satisfy is crucial for ensuring reliable water supply and planning expansions. This maximum is computed by solving a mixed-integer nonlinear program (MINLP), which is computationally challenging. The challenge is typically addressed by solving mixed-integer linear program (MILP) relaxations obtained through variable domain partitioning with linear over- and under-estimators of the partitions. These MILPs provide useful bounds and are relatively easier to solve. However their solutions can violate nonlinear water flow physics. Thus, recovering feasible MINLP solutions from the MILP relaxations is crucial for enhancing MILP-based approaches.

In this presentation, we propose a robust solution recovery method that efficiently computes feasible MINLP solutions from MILP relaxations, regardless of partition granularity. Together with iterative partition refinement, our method produces feasible solutions that progressively approach the optimal solution. Through extensive numerical experiments, we demonstrate that the proposed method outperforms other recovery methods and direct MINLP solves by consistently recovering high-quality feasible solutions with significantly reduced computation times.

17 - Personalized Pathways to Discharge: Explaining What Matters, for Whom, and Why

Ha Na Cho, UC Irvine, Irvine, CA, United States, Kai Zheng, Bow Hansen, Alexander Lopez, Sairam Sutari

Efficient discharge planning is a critical challenge in surgical care operations, particularly for optimizing bed turnover and care coordination. We present a predictive framework to identify whether a patient who underwent surgery is likely to be discharged the following day, based on unstructured clinical notes. Leveraging data from a surgery department, we evaluated multiple machine learning (ML) models, including Logistic Regression, Random Forest, XGBoost, Support Vector Machines, and Naive Bayes, under class-balanced settings to address outcome imbalance. Among the models, XGBoost yielded the highest performance (accuracy: 0.81), followed closely by Random Forest and SVM (0.74-0.81). Logistic Regression demonstrated high recall (0.72) in detecting next-day discharges but exhibited lower precision.

To ensure transparency, we employed SHAP (SHapley Additive exPlanations) to analyze both global and patient-level feature attributions. SHAP summary plots revealed that certain textual features such as "home today" and "no acute distress" contributed positively to next-day discharge classification, while terms like "dressing change" and "dependent total" were more influential in predicting prolonged hospitalization. These results are not causal but offer interpretable operational proxies for discharge readiness.

At the individual level, SHAP waterfall plots highlighted significant heterogeneity in influential features, confirming that discharge decisions are highly personalized. Different features carried predictive weight for different patients, reflecting context-dependent pathways to discharge even within the same outcome class.

This study emphasizes the clinical value of explainable ML in supporting anticipatory discharge decisions. Our approach enables clinicians to trace the reasoning behind next-day discharge predictions, enhancing trust, and efficient resource planning.

18 - Navigating the Digital Footprint: a Cross-Cultural Analysis of Parental Information Disclosure on Social Media Platforms

Paras Bhatt, University of Alabama in Huntsville, Huntsville, AL, United States, Tripti Singh

The widespread practice of “sharenting,” where parents share their children’s information on social media, raises critical questions about how platform design and cultural norms jointly shape digital footprints and privacy risks. Although prior research has explored parental motivations for sharenting and examined individual platform features, such as content persistence (long-lasting, easily retrievable content) and ephemerality (temporary, short-lived content), a critical gap remains in understanding how combinations of platform-level affordances (e.g., algorithmic curation, privacy controls) interact with culturally embedded privacy norms to shape sharenting and associated privacy risks.

Using a mixed-methods study informed by affordance theory and social processing theory, we first map the platform’s key affordances, such as visibility, content persistence, editability, interactivity, algorithmic curation, and data governance, to analyze publicly available sharenting posts and demonstrate how these affordances shape parental disclosure. This includes a cross-platform analysis (YouTube, TikTok, and Instagram) across two different cultures: an individualistic, low-context culture (e.g., United States) that values direct and explicit communication, and a collectivist, high-context culture (e.g., Nepal), which emphasizes indirect cues, shared social norms, and group harmony. Next, we conduct a survey of parents in both cultures to assess how affordance-culture interactions drive parents’ oversharing versus protective behaviors.

Our findings will demonstrate 1) specific platform features and culture pairings that amplify children’s exposure and privacy risks, (2) reveal how cultural privacy norms moderate parental use of platform controls, and (3) generate actionable guidelines for designers to tailor affordances that encourage mindful sharing for safeguarding child privacy in diverse digital ecosystems.

19 - A Data-Driven Evaluation of Primary Care Utilization Across Pandemic Stages

Jennifer Mendoza-Alonzo, Department of Statistics, Franklin College of Arts and Sciences, University of Georgia, Athens, GA, United States

Primary care services are essential for preventive health and managing chronic conditions, influencing overall population health outcomes. Social determinants of health significantly influence individuals’ access to and utilization of primary care services. This cross-sectional study evaluates changes in primary care utilization across pre-, peak-, and post-pandemic periods by applying machine learning and traditional methods. Using Behavioral Risk Factor Surveillance System (BRFSS) data, we assess risk factors for adults receiving routine care based on social determinants of health and chronic conditions. We examine variable importance using machine learning and SHAP (SHapley Additive exPlanations), and incorporate the most relevant variables into traditional statistical models. Our findings offer insights into healthcare access patterns and demonstrate the value of combining interpretable machine learning algorithms and public health analytics to inform equitable health policy and strengthen health system resilience.

20 - Route Optimization for Electric Trucks in Solid Waste Collection: Impacts of Network Representations and New Solution Algorithms

Chinonso Okorie, FAMU-FSU College of Engineering, Tallahassee, FL, United States, Yanshuo Sun

Solid waste collection (SWC) is vital for environmental sustainability and public health. With the adoption of electric trucks (ETs), recent studies have begun to address ET routing optimization, but none has considered the realistic transportation network on which SWC occurs. This study bridges that gap by comparing two network representations: a real-world network with highway nodes and links, and an alternative network with virtual links. Comparative analysis shows the real-world network dominates when the network is sparse and most communities are visited, while the alternative excels when highways require multiple visits. Additionally, we propose two solution algorithms, namely column generation (CG) and sequential decomposition heuristic (SDH) to solve large-scale problems. Extensive computational experiments on synthetic and real-world case studies show their effectiveness in reducing computation time by up to 90%.

21 - Spatiotemporal Endemic&Ndash;Epidemic Modeling of COVID-19 Using Cluster-Aggregated Mobility Data During a Natural Disaster

Davison Tsekeni, Florida A&M University, Florida State University (FAMU-FSU) College of Engineering, Tallahassee, FL , Tallahassee, FL, United States, Arda Vanli

Understanding how mobility patterns influence infectious disease dynamics is critical for modeling epidemic spread, particularly during overlapping crises. We present an endemic–epidemic (EE) modeling framework by clustering ZIP codes using a k-means approach on geographic centroids and aggregating both COVID-19 infection case counts and origin-destination mobility data. Our study focuses on Florida’s Escambia, Santa Rosa, and Okaloosa counties from September 1 to September 30, 2020, encompassing the landfall of Hurricane Sally. To capture behavior-driven changes in mobility, we model inter-cluster movement rates using a multi-step changepoint approach. EE models are fit with distributed lag structures and Poisson likelihoods, and are evaluated using AIC, R^2 , and Residual Sum of Squares (RSS). The model estimates region-specific (cluster-specific) reproductive numbers and decomposes infection dynamics into self-driven, neighbor-induced, and environmental components. Results show that cluster-level variability in outbound mobility is strongly associated with neighbor-driven infection contributions. This modeling approach supports operational decision-making during public health emergencies by enabling local health agencies to identify high-risk regions, anticipate outbreak trajectories, and evaluate mobility-targeted interventions. By integrating real-world mobility disruptions into a scalable forecasting framework, this study contributes a practical tool for epidemic risk assessment, particularly in settings affected by concurrent public health and environmental hazards.

22 - A Two-Stage Stochastic Optimization Model for Equitable and Cost-Effective Food Distribution Under Multiple Uncertainties: A Case Study of Hays County Food Bank

Mukitil Islam, Texas State University, San Marcos, TX, United States, Clara Novoa

This poster presents a two-stage stochastic optimization model developed to support strategic and operational decision-making at the Hays County Food Bank (HCFB) located in Central Texas. The model incorporates practical complexities faced by the HCFB, including perishability of food, storage capacity limitations, restricted transportation resources, and uncertainty in donations, volunteers’ availability, and demands from food-insecure customers.

The model aims to minimize the total annual cost of food distribution while ensuring a weekly equitable distribution in six food categories and four cities. The model is solved using two methods, assuming discrete scenarios and employing sample-average approximation. The numerical results provided by these methods are validated against data and key performance indicators provided by HCFB. Through experimentation with the model, it has been found that the model is effective in assisting HCFB with future expansion plans, enhancing accessibility for vulnerable populations, and reducing food waste.

23 - Testing Optimization Models for Production

Francisco Zenteno Smith, American Airlines, Forth Worth, TX, United States, Jonathan Yu

Optimization models are deeply embedded in modern operational systems, powering decisions that move goods and people across the globe. In the commercial aviation industry in particular, these models play a central role in long range planning, day-to-day operations, and on-demand disruption recovery. For example, inclement weather in major hubs can cause cancellations that ripple across the network. Dependable and resilient models are the deciding factor in millions of dollars in savings, or loss.

While optimization models are mainly mathematical constructs, writing, deploying, and maintaining them is inherently a software engineering challenge. As business

needs evolve, so must the model, in lock-step with the code.

In this work we present a framework developed at American Airlines that enables large teams to collaboratively build and maintain optimization models efficiently, without compromising resiliency. The framework integrates continuous integration (CI) pipelines to automate both unit and integration testing of optimization models. Integration tests validate complete model snapshots against expected formulations, while unit tests verify individual model components (such as constraints, variables, and objective function terms) in isolation. This structured approach ensures alignment with business requirements, detects regressions early, and significantly reduces the need for time-consuming manual testing.

We show how this approach supports early detection of bugs, reduces deployment risk, and enhances trust in the optimization tools that support operational decision-making. By embedding rigorous testing into the development lifecycle, our framework improves reliability and accelerates the delivery of optimization-based solutions in production environments.

24 - An Adaptive Decision-Making and Environmental Monitoring System for Collaborative Robots Using Sensors

mahrou samizadeh, auburn university, auburn, AL, United States, Konstantinos Mykoniatis

As collaborative robots (cobots) become more prevalent in manufacturing, logistics, and service industries, their ability to adapt to dynamic environments is critical. Traditional cobots typically operate based on pre-programmed rules, which limit their flexibility and responsiveness in real-world conditions. This project presents a two-phase approach to improving cobot intelligence through real-time environmental sensing and decision-making.

In Phase 1, we developed a rule-based control system that integrates ultrasonic, infrared (IR), and DHT11 (temperature and humidity) sensors. These sensors were connected to an Arduino microcontroller, which processed sensor data and communicated binary signals to a Universal Robot (UR) arm. The robot responded in real time to object presence, executing pick-and-place actions. We evaluated sensor configurations under varying environmental conditions, and results from 40–50 experimental trials showed that ultrasonic sensors provided more stable and faster responses compared to IR sensors. ANOVA analysis confirmed the significance of sensor choice on robot performance, establishing a benchmark for real-time detection accuracy and task efficiency.

In Phase 2 (under development), the same sensor inputs will be combined with computer vision using YOLOv8 and a machine learning model to drive adaptive cobot behavior. The ML model will analyze multi-sensor patterns and YOLO detections to make context-aware decisions without relying on hard-coded rules.

This work contributes to the growing field of Human-Robot Interaction by proposing a scalable framework for building intelligent cobots. It supports the Industry 5.0 vision of human-centered automation by enabling cobots to make faster, smarter decisions in complex, changing environments.

25 - Convexification of Multi-Period Quadratic Programs with Indicators

Jisun Lee, UC Berkeley, Berkeley, CA, United States, Alper Atamturk, Andres Gomez

We consider a multi-period convex quadratic optimization problem with indicator variables, where the state evolves dynamically as an affine function of the current state and input (control) variables in each period. The affine function is determined by the indicator variables that denote the mode chosen in each period. This problem class has important applications in hybrid control and statistical learning. We study the complexity of the problem in various settings, demonstrating both polynomial-time solvable and NP-hard cases. For the polynomial-time solvable case, we present a compact convex hull description in an extended space with linear and conic quadratic inequalities. We also propose a polynomial-time algorithm that solves the problem as a shortest path problem for the unbounded case. Computational experiments with data from calcium imaging data deconvolution and hybrid-electric vehicle power management indicate promises and challenges.

26 - Modeling Mental Health Delivery via a Decentralized System

Candace Yano, UC Berkeley, Berkeley, CA, United States, Ricky Huang

Regional mental health services are often provided by a partially decentralized and capacity-constrained network of organizations operated by various government agencies and nonprofit entities, with the latter being supported by some combination of government subsidies, charitable contributions, and volunteer labor. In this setting, quantifying how the funding level or policy changes at one entity affects the overall value provided by the system in terms of improved health of its clients is difficult. We report on the initial development of a model to aid in understanding these relationships based on the underlying situation in one U.S. county.

27 - Inter-Airline Cooperative Disruption Management with TIME-Varying Cost Functions

Tolga Atabas, University of Michigan, Ann Arbor, MI, United States, Max Li, Brian Denton

Airlines regularly face delays due to disruptions that constrain airport capacities, such as adverse weather events. In response to these disruptive events, air traffic managers reschedule flights in order to balance the constrained airport capacities with the incoming demand of departing and arriving flights, leading to many flights being delayed. Airlines can create a cost of delay function that models the cost (e.g., dollars) for each flight as a function of the severity of delay (e.g., delay hours). These functions are then used in the optimization models to determine the optimal slot swaps that minimize costs for airlines. Recent methods use intra- and inter-airline slot swapping to reduce delay costs. However, the existing approach assumes fixed-rate cost functions that fail to capture real-world cost surges associated with missed connections, crew disruptions, and tail-propagated delays (i.e., delays for subsequent flights performed by the same aircraft). This study develops a novel, convex, stochastically-generated piecewise linear cost function that more accurately reflects time-varying delay costs. Using a Poisson process to model the times for cost rate surge events, we integrate these cost functions into the DELED framework and construct a Multi-Integer Linear Programming (MILP) optimization model. This enhancement improves the realism and robustness of airline delay management models, offering new insights into operational slot swap decision-making. We will show that incorporating time-varying cost structures can improve optimal delay assignments compared to fixed-rate cost assumptions.

28 - Enabling Collaborative Vehicle Routing by Mechanism Design

Jack Rissman, Northern Illinois University, Dekalb, IL, United States, Ziteng Wang

A fundamental question when looking at collaborative vehicle routing is whether the participating logistics providers will benefit from the coalition. These benefits can be in the form of cost reduction, efficiency improvements, environmental benefits, enhanced service levels, and room for innovation. This research aims to design an efficient mechanism for allocating cost savings in order to incentivize logistics providers to join the grand coalition. The research began with revealing Shapley value-based allocation and proportional allocation are not efficient because the core of the underlying cooperative game could be empty. Therefore, some outside source of support is necessary to enable the grand coalition. The next steps consist of optimizing the slack of each allocation savings to find a minimal value of necessary outside support. The research aims to find a mechanism to decide the outside support to better incentivize the collaboration of stakeholders. We will then generalize the results to collaborative vehicle routing problems involving any number of logistics providers.

29 - Average Distance of Random Bipartite Matching in Networks

Yuhui Zhai, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Shiyu Shen, Yanfeng Ouyang

The bipartite matching problem is widely applied in the field of transportation; e.g., to find optimal matches between supply and demand over time and space. Recent efforts have been made on developing analytical formulas to estimate the expected matching distance in bipartite matching with randomly distributed vertices in two- or higher-dimensional spaces, but no accurate formulas currently exist for one-dimensional problems. This paper presents a set of closed-form formulas, without curve-fitting, that can provide accurate average distance estimates for one-dimensional random bipartite matching problems (RBMP). We first focus on one-dimensional space and propose a new method that relates the corresponding matching distance to the area size between a random walk path and the x-axis. This result directly leads to a straightforward closed-form formula for balanced RBMPs. For unbalanced RBMPs, we first analyze the properties of an unbalanced random walk that can be related to balanced RBMPs after optimally removing a subset of unmatched points, and then derive a set of approximate formulas. Additionally, we build upon an optimal point

removal strategy to derive a set of recursive formulas that can provide more accurate estimates. Then, we shift our focus to regular discrete networks, and use the one-dimensional results as building blocks to derive RBMP formulas. To verify the accuracy of the proposed formulas, a set of Monte-Carlo simulations are generated for a variety of matching problems settings. Results indicate that our proposed formulas provide quite accurate distance estimations for one-dimensional line segments and discrete networks under a variety of conditions.

30 - Fitness Center Operational Improvements with Discrete Event Simulation

Junayed Pasha, Gannon University, Erie, PA, United States, Chenay Date Line, Jayed Pasha, Emilio Bell, Andres Esteban Florez

Fitness centers often deal with a large number of users. This may lead to frequent breakdowns of equipment and reduce the operational efficiency of the entire fitness center. Therefore, this research aims to analyze and improve the operations of a fitness center. Considering user flow and equipment utilization, this study seeks to identify bottlenecks, reduce wait times, and improve the overall user experience. In order to replicate the fitness center's operations, a simulation model is developed, which incorporates various features, such as user arrival patterns, equipment usage and popularity, queuing systems for high-demand equipment, as well as equipment reliability and downtime. Relevant data on usage, equipment layout, and failure rates is also collected. In order to improve the fitness center's operations, the simulation explores various scenarios (e.g., equipment rearrangement, maintenance strategies). Simulation results provide insights into peak usage times and the associated congestion, overall equipment utilization rates, impact of breakdowns on overall efficiency, and effectiveness of different layout configurations. Based on the simulation outcomes, this study provides some recommendations, including a redesigned equipment layout to reduce congestion, creating different strategies for managing high-demand machines, a modified maintenance schedule to minimize downtime, and resource allocation suggestions for popular equipment. These data-driven recommendations aim to enhance the fitness center's operational efficiency and user satisfaction, supporting informed decision-making for facility management.

31 - A Feature-Enhanced Capacity Prediction Method for Efficient Lithium-Ion Cell Grading Under Multi-Temperature Conditions

Shih Han Huang, Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, United States, Kwok Tsui, Xinwei Deng

In battery pack assembly, cells are graded by measuring discharge capacity to enable consistent module performance. However, traditional grading tests often consider the capacity at a single temperature, which does not reflect the diverse environmental conditions for battery packs in real-world use. In this work, we present a data-driven grading method by utilizing features from the charging phase and the first half of the discharge phase to predict full discharge capacities under multiple temperature conditions. The proposed method enables early-stage prediction and supports grading based on the multi-temperature predictions and their uncertainties. Compared to conventional approaches, the proposed method reduces testing time and cost while improving grading robustness. The proposed method is validated across three major lithium-ion chemistries to demonstrate its merits. It also offers practical potential for quality control in battery pack manufacturing.

32 - Toward Smarter U-Net Architectures: A Systematic Review of Optimization Strategies for Tumor Segmentation

Omar Abueed, Binghamton University, BINGHAMTON, NY, United States, Yong Wang

Since its introduction in 2015, the U-Net architecture has become a foundational model in medical image segmentation, particularly for tumor delineation. However, challenges such as tumor heterogeneity, blurred boundaries, and partial volume effects continue to hinder precise segmentation outcomes. To address these limitations, a wide array of optimization strategies have been proposed to enhance U-Net's performance in segmenting CT and MRI images. This systematic review, conducted in alignment with the PRISMA guidelines, categorizes and critically analyzes the most recent advancements in U-Net optimization across seven major domains: backbone enhancements, bottleneck refinements, skip connection strategies, Transformer-based modules, metaheuristic algorithms, state space models (SSMs), and self-adaptive architectures. The review emphasizes how each technique contributes to overcoming the inherent challenges of the original U-Net design. Additionally, the significance of preprocessing methodologies in boosting segmentation effectiveness is thoroughly explored. This review not only synthesizes current optimization trends but also highlights persisting research gaps, offering valuable directions for future studies aiming to enhance U-Net's accuracy, robustness, and clinical utility in tumor segmentation tasks.

33 - Exploring Trust Dynamics in Stag-Hunt Games: a Simulation-Based Study

Mobin Zarreh, Arizona State University, Tempe, AZ, United States, Paul Grogan

From resource management to autonomous supply chains and other distributed systems, effective collaboration depends on trust among autonomous agents. These settings involve complex decision-making under uncertainty, where stakeholders must assess risks and anticipate others' actions, making them well-suited for decision analysis and game-theoretic modeling. Understanding how trust evolves and deteriorates can inform policy and the design of resilient multi-agent systems. This study introduces a novel dual-state agent-based model within the Stag Hunt game framework to investigate how trust evolves and shapes decision-making across diverse partner strategies. Unlike traditional scalar trust models, our framework explicitly separates two internal states: signal (the agent's perceived probability of partner cooperation) and trust (confidence in that expectation). This dual-state structure enables distinct, nuanced updating mechanisms. Trust updates depend on alignment between expected and observed partner behavior, whereas the signal evolves from observed collaboration frequencies weighted by recency. We evaluate the model against a range of partner strategies, including cooperative, adversarial, and particularly deceptive opponents such as an expectation-violation cheater who intentionally acts contrary to the agent's expectations. Results show that consistent cooperation fosters sustained collaboration, while predictable defection, despite building trust and confidence, leads to independent choices. Deceptive strategies notably erode trust, significantly reducing cooperation. Our findings highlight how expectation alignment, predictability, and learning dynamics shape strategic outcomes, offering valuable contributions to behavioral game theory and guiding the design of resilient multi-agent collaborative systems.

34 - Reconstructing and Predicting Driver's Routing Behaviour Under Incidents with Large Language Models

Haocheng Duan, Carnegie Mellon University, Pittsburgh, PA, United States, Sean Qian

Accurately forecasting network-wide impacts of traffic incidents demands a precise grasp of how drivers re-route when disruptions occur. Yet route choice emerges from a chain of context-dependent decisions that grow even more intricate once incident information and nonrecurrent network conditions enter the mix. Complicating matters, GPS traces that reveal these choices are both massive and sparsely sampled, making direct interpretation labor-intensive. To tackle these challenges, we employ a large language model (LLM) as an autonomous agent to preprocess extensive GPS datasets alongside a large-scale road network, obtain possible route information between, and complete trace subsets affected by disruptions. We then fine-tune a decision-making LLM on the small, carefully curated subset using supervised fine-tuning (SFT) and reinforcement learning from human feedback (RLHF), aligning its reasoning with human routing logic. By testing on real-world data collected in the Washington DC region in 2020, we found this two-stage pipeline substantially improves the model's ability to capture drivers' re-routing behavior and significantly boosts the accuracy of route-choice predictions during traffic incidents.

35 - Using Simple Math to Quantify the Impact of the Triangle Trade and Enslavement Practices on Familial Lines: A Case Study of Colonial Ayiti

Jovan Julien, Georgia Institute of Technology, Atlanta, GA, United States

Topic:

The impact of colonialism and enslavement practices on genetic diversity, cultural survival, and mortality in colonial Santo Domingue is not well documented. Due to the paucity of data on the enslaved population, mathematical modeling can serve as an effective way to explore the impact of these brutal regimes.

We leverage mathematical methods to quantify the impact of imperialism and colonialism between 1492 and 1804.

Methodology

Our goal is to explore the if(s). «Yo pa fe listwa avek les si» an memn tan jistis depandan sou « si ». This quantification of ifs is essential for justice, for the qualification of reparation and restitution, and to mark a historical tragedy such that generations might avoid its replication

We apply a semi-markov model framework to replicate and understand the creation process of a new population, Haitians, during the colonial period. Utilizing historical patterns of immigration and mortality we fit a model for family line and cultural community mortality and survival. Counterfactual analyses are conducted to quantify the impact if a revolution was not started and if French attempts to recapture the island had been successful.

Key arguments and results

Leveraging a land-based analysis allows for quantification of the crimes that occurred centuries ago. We offer a framework by which interested parties can explore the impact of colonialism.

Conclusion

While many would argue that mathematical models cannot capture the dreams unrealized, we offer a method by which a value can be generated for the many generations lost to Ayiti since colonization.

36 - Computer-Aided Analysis of Decentralized Optimization Algorithms

Ning Fang, Cornell University, Ithaca, NY, United States, Xin Jiang

Performance Estimation Problem (PEP) theory offers a rigorous framework to analyze the worst-case convergence rates of first-order optimization algorithms. We extend PEP to the decentralized optimization setting by developing a generalized primal formulation that accommodates both static and time-varying communication graphs, enabling systematic analysis of decentralized algorithms across different network scenarios. By encoding decentralized updates using Gram matrix representations, we systematically capture interpolation constraints involving function values, gradients, and consensus violations across agents. We apply this framework to analyze various decentralized gradient-based methods and present some intriguing results that are unlikely to be discovered without computer-aided analysis.

37 - Data-Driven Insights from Raccoon Eyes: Quantifying the Financial Impact of AI-Powered Food Waste Reduction at Georgia State University

Izuhunwa Amadasun, Georgia State University, Atlanta, GA, United States, Jennifer Wilson, Ivan Zou, Adetoun Adeleke, Mahira Begum, Lauretta Fagbuyi, Alexis Mathis, Tobi Oyeniran, Davron Dean, Chidimma Ihejirika, Wendy Roth, Nesreen El-Rayes

Food waste in institutional dining environments poses a significant financial and environmental challenge. This case study from Georgia State University's Panther Dining explores how the Raccoon Eyes AI dashboard (a computer vision and analytics platform) can drive data-informed decisions to reduce waste and associated costs. Drawing on over one million meal records, 42,000 pounds of waste data, and detailed breakdowns by food type, time, and plate count, the study identifies high-waste items and operational inefficiencies. By monitoring patterns such as overproduction and plate returns, and implementing targeted interventions (e.g., portion resizing, utensil adjustments, and menu optimization), the research quantifies both waste reduction and cost savings. Time series and intervention analyses demonstrate the platform's ability to guide waste mitigation strategies that improve efficiency and sustainability. The findings offer a replicable framework for other institutions to leverage AI tools like Raccoon Eyes to improve operational performance and reduce the financial burden of food waste. Key questions addressed include, but are not limited to: Which food items contribute most to waste and cost? What measurable savings result from intervention strategies? The results underscore the economic justification for adopting AI-powered waste tracking, highlighting its dual impact on institutional budgets and environmental outcomes.

38 - Mip Models for Searching---Bayesian Updates

Ransheng Guan, University of Wisconsin-Madison, Madison, WI, United States, Jeff Linderorth

Anti-Submarine Warfare (ASW) operations necessitate the strategic allocation of limited, often diverse, assets to locate elusive underwater threats. Foundational search theories leverage all information, including vital 'negative search evidence'—the absence of detection—to recursively refine probabilistic estimates of target locations using Bayesian inference. However, a primary challenge in dynamic operational environments is that the intelligence value and outcomes of future searches are not predetermined. Instead, they are significantly influenced by prior asset deployments and their results. This introduces complexities associated with Decision-Dependent Uncertainty (DDU), where actions actively shape the uncertainties faced in subsequent stages.

We introduce a mathematical programming framework to tackle ASW search problems where the information landscape evolves based on operational decisions. This approach combines established belief updating techniques with principles for handling DDU. It aims to formally model how search actions and their consequences dynamically alter the probabilistic assessments guiding future efforts. Recognizing that capturing these dynamic, often non-linear, interactions can lead to complex optimization challenges, this framework provides a structured methodology for incorporating decision-dependent information, including the impact of negative evidence, into ASW planning over a defined operational theater and time horizon.

39 - Harnessing Sentiment Analysis for Sustainable Dining: An AI-Driven Tool to Reduce Food Waste

Nesreen El-Rayes, Georgia State University, Atlanta, GA, United States, Jennifer Wilson, Ivan Zou, Izuhunwa Amadasun, Victoria Kehinde, Juliet Donkor, Alexis Mathis, Tobi Oyeniran, Wendy Roth

This study explores the role of consumer sentiment in shaping food waste outcomes within institutional dining environments, focusing on how qualitative feedback can inform sustainable decision-making. Leveraging a vision AI platform (Raccoon Eyes) that simultaneously tracks food waste metrics and captures students' feedback in the dining hall at Georgia State University, we apply natural language processing (NLP) and sentiment analysis to thousands of reviews collected across campus dining halls.

Integrating sentiment data with quantitative indicators such as waste volume, plate counts, and cost, the analysis highlights how AI-enabled tools can enhance responsiveness and sustainability in large-scale food service operations. By analyzing the sentiment polarity with waste patterns for specific food items, the research examines the sentiment as a leading indicator of high-waste meals. Additionally, we assess the impact of feedback-informed interventions on subsequent waste reduction and cost efficiency.

This research contributes to the growing intersection of AI, sustainability analytics, and operational efficiency in the food service sector, and has managerial implications that can be cascaded in similar setup in other organizations.

40 - Koopman Representations with Irregular TIME Intervals

Younghwan Cho, University of Illinois Urbana-Champaign, Champaign, IL, United States, Richard Sowers

Koopman operator theory has been widely applied in data assimilation problems of real systems governed by dynamical systems, as the theory allows for data-driven construction of modes of dynamical systems. In many modern problems, Koopman modes may have to be learned from data with irregular sampling intervals, as opposed to commonly used regular sampled data. We here propose a framework to recover Koopman eigenfunctions and eigenvalues for irregularly sampled data. We show that a Koopman eigenfunction and eigenvalue can be recovered via a natural optimization problem. We also provide technical remarks on the anticipated challenges in optimization and suggest a procedure to address them. Simulation studies under different irregular sampling scenarios verify the robustness of the proposed method in learning Koopman eigenfunctions.

41 - Exploring the Correlation Between Winter Severity Indices and Winter Maintenance Costs: Insights and Comparative Analysis from Nationwide Practices

Ardeshir Fadaei, Michigan State University, East Lansing, MI, United States, Farish Jazlan, Amirali Soltanpour, Hamid Mozafari, Ali Zockaie, Mehrnaz Ghamami, James Roath

Winter maintenance operations are essential for safe and efficient transportation, yet unpredictable weather introduces major challenges in cost and material planning. A single severe snowfall can deplete resources midway through the season, while milder-than-expected winters may result in unused surpluses. The Winter Severity Index provides a standardized measure of winter intensity, offering a promising tool to address these challenges. This study explores the relationship between WSIs and winter maintenance costs through a nationwide survey of state DOTs and Clear Roads Survey data. The Clear Roads survey analysis involves regression models, assessing the relationship between the AWSSI and winter maintenance expenditures across different U.S. climate divisions. AWSSI was used not as a recommended tool for operations, but as a nationally available and consistent index to investigate whether winter severity is statistically associated with cost trends at a broad scale. The study also assessed

non-linear regression models, which yielded similar results to linear models; therefore, the linear model was retained for its simplicity and interpretability. Results showcase significant correlation within various climate clusters, indicating that weather severity can explain cost variability. The study reveals that WSIs can improve winter maintenance budget management, but their effectiveness is regionally nuanced. Customized WSIs are essential for better cost and resource planning, as no single index, including AWSSI, works optimally in all geographic contexts. The study also identified best practices by state DOTs, such as integrating WSIs with real-time tools like Road Weather Information Service and Maintenance Decision Support System, to enhance planning accuracy and responsiveness.

42 - A Rose Given, Fragrance in Hand: Impact of Prize Draw Variation on Sales Change

Larry Cui, Portola High School, Irvine, CA, United States

Digital technologies are increasingly being adopted by companies to enhance various aspects of performance, including internal operations, financial management, marketing, and sales. This study utilizes prize draw and sales data from a publicly listed company that implemented digital technology to conduct consumer prize draws. By analyzing 8,872 observations across multiple cities in a large country, the study examines how variations in prize draw amounts influence changes in sales. The findings suggest that fluctuations in prize draw incentives can have a significant impact on product sales.

43 - Enhancing Energy Resilience of Data Centers Through Flexible-Load Strategies

Mohamed Abdelhady, Department of Multidisciplinary Engineering, Texas A&M University, College station, TX, United States, Eleftherios Iakovou, Stratos Pistikopoulos

Hyperscale data centers may face mounting reliability and cost challenges as power grids struggle to expand capacity under a volatile, uncertain, complex, and ambiguous (VUCA) environment. Traditional grid procurement strategies may no longer provide adequate resilience or cost-effectiveness in these challenging conditions. This study evaluates whether data centers should transition solely to reliance on the grid to integrating on-site self-generation portfolios that combine photovoltaic (PV), wind, gas turbines, nuclear, and battery energy storage systems (BESS).

We develop a two-stage stochastic programming framework to identify optimal technology portfolios under uncertainty. Our approach uses a scenario-generation algorithm that synthesizes fourteen years of historical wholesale price data, preserving critical patterns including seasonality and volatility. The first stage determines optimal capacity investments across technologies, while the second stage optimizes dispatch decisions to minimize expected costs while providing grid flexibility services. These services not only capture value from negative-price events and peak-shifting arbitrage but also provide rapid load support to relieve system stress and prevent grid blackouts. Under baseline conditions, PV and BESS emerge as the cost-optimal combination: PV panels provide low-marginal-cost energy during daylight hours, while BESS shifts energy to peak periods and captures value from negative-price events. However, when we incorporate explicit reliability constraints through chance-constrained programming—capping the probability of unmet load below specified thresholds—gas turbines re-enter the optimal mix to hedge against extreme price spikes and supply shortfalls from the grid.

These findings may offer data center operators and policymakers clear guidance on developing robust energy portfolios under power grid uncertainty.

44 - Optimizing and Deploying Schematic Bicycle Maps with Milp and User Input

Mohammed Abdalazeem, University of Massachusetts Amherst, Amherst, MA, United States, Jimi Oke

Schematic maps simplify network complexity and improve cognitive accessibility, principles widely adopted in public transit but rarely applied to bicycle infrastructure. This project proposes a complete pipeline for the development of schematic bicycle maps, including design rule synthesis, user-centered input via surveys, and an optimization-based automated layout engine. We begin by extracting design principles from existing bicycle maps and transit schematics, then synthesize a new set of constraints specific to cycling, accounting for elevation, stress level, infrastructure type, and the absence of fixed lines. We formalize these principles into a mixed-integer linear programming (MILP) model to generate schematic layouts. The optimization minimizes distortion and visual complexity while adhering to derived cyclist-centered design rules. To validate the usability of the proposed approach, we conduct a two-phase survey in the case study area to capture user preferences before and after exposure to generated schematics. We also develop a bicycle demand model using public and sensor-based count data, enabling future impact evaluation. The final map will be selected through focus group feedback and prepared for public installation. This work offers a replicable framework for schematic bike maps and a planning tool for cities aiming to expand cycling for sustainable, healthy, and accessible mobility.

45 - Fourier Learning Machines and Their Applications

Gabriel Nicolosi, Missouri University of Science and Technology, Rolla, MO, United States, Mominul Rubel, Adam Meyers

This work is concerned with the historical, computational and mathematical relationships between classic Fourier Series and Neural Networks. For five decades, numerous attempts to unify these two mathematical tools have been made by researchers in the most diverse yet often disconnected fields. In this work, a historical perspective on this endeavor is presented, alongside a new, straightforward and computationally efficient representation of Multivariable Finite Fourier Series as a Multi Layer Perceptron, referred to as a Fourier Learning Machine. Furthermore, we study a variation wherein integrality constraints over the frequencies are relaxed, and its relations to the Fourier Transform are investigated. Computational experiments and comparative performance evaluations in the context of Scientific Machine Learning and Numerical Optimal Control are presented.

46 - Why Do College Students Cheat?

Matt Dean, William & Mary, Williamsburg, VA, United States

Academic dishonesty among college students is commonplace. Building off a well-known article in the *Journal of Business Ethics* (Simkin and McLeod 2010), our study describes the status of college students cheating in 2024. We surveyed 171 business school students and analyzed their responses by applying partial least squares methodology on cheating behavior. Among cheaters in the previous 2010 study, a “desire to get ahead” was the most important motivating factor. Our study went deeper into this phenomenon to understand why college students are cheating by engaging with a focus group of survey participants. The results better describe when students choose to cheat and what professors and administrators should do to prevent such behavior.

47 - A Mitigation Strategy Against Cyberattacks on Distributed Optimal Power Flow

Mohannad Alkhrajah, National Renewable Energy Laboratory, Golden, CO, United States, Daniel Molzahn

Distributed optimization offers desirable features for power system operations by enabling coordination among multiple agents with minimal data exchange. However, distributed optimization's inherent reliance on communications among multiple agents potentially increases vulnerabilities to cyberattacks and communication failures. Leveraging concepts from robust optimization, this paper presents a mitigation strategy to reduce the impacts of cyberattacks and communication failures on distributed optimization for power systems. The proposed mitigation strategy involves a distributed solution of a bilevel *admissible bounds* problem. This problem bounds the maximum flows that can be achieved in branches connected to compromised areas without violating the local operational constraints of other areas. To solve this problem, this paper first introduces a decomposition method based on Kron reduction that uses a Power Transfer Distribution Factor (PTDF) formulation. The proposed method then reformulates the bilevel admissible bound problem using the weak duality theorem to reduce the problem into a single level. This paper demonstrates the proposed method using several test systems and compares the results with the global central solution. The results show the trade-off between solution quality and computation time when solving the admissible bounds problem.

48 - Improving Work Zone Safety: Assessing Law Enforcement Effectiveness with a Split-Plot Design in South Carolina

Mahyar Madarshahian, University of nebraska lincoln, Lincoln, NE, United States, Reka Howard, Nathan Huynh, Jackson Wegmet, Chowdhury Siddiqui

This study evaluates the impact of law enforcement presence on vehicle speeds in South Carolina work zones using a split-plot design with blocking, which offers greater precision compared to traditional factorial designs. Utilizing 2019 traffic data, eight alternative models are developed, examining two response variables: average speed

across the entire work zone and in the transition area. Each response variable is analyzed in two ways: (1) by calculating the difference between average speed and the temporary posted speed limit to assess compliance, and (2) by including traffic volume as a covariate to improve model accuracy. The findings reveal that the presence of state troopers significantly reduces vehicle speeds and enhances compliance with temporary speed limits in both zones. Seasonal variation was generally not a significant factor, with no effect observed in six of the eight models. Importantly, incorporating traffic volume as a covariate improves model precision, supporting its inclusion in future analyses.

49 - Keypoint-Aware Human Body Part Segmentation via a Dual-Discriminator Generative Adversarial Network

Shivam Shrivastava, University of Missouri, Columbia, MO, United States, Hyeong Suk Na

Accurate human pose estimation and body part segmentation are essential for various applications, including sports analytics, video surveillance, and security systems. While these two tasks have traditionally been addressed independently, they are inherently complementary and can be improved by joint learning. Existing studies mainly focus on leveraging segmentation to enhance pose estimation; however, very few models have exploited human pose keypoints to improve body part segmentation. To consider this gap, we develop a cross-domain learning framework to improve human body part segmentation by utilizing keypoints obtained from human pose estimation. Our model architecture is a dual-discriminator generative adversarial network, with one discriminator dedicated to keypoint detection and the other to body part segmentation. The generator has a shared convolutional backbone, followed by two task-specific branches. While the keypoint detection branch employs only the channel-attention module for keypoint regression, the segmentation branch integrates both channel-attention and spatial-attention modules. The channel-attention module is designed to enhance the regression of keypoint heatmaps across binary channels corresponding to each body joint. Additionally, the spatial-attention module aggregates concatenated features from the channel-attention module and the shared convolutional backbone to enhance spatial localization in segmentation. Then dual discriminators are employed to train these two modules individually, resulting in more accurate segmentation boundaries. The model is evaluated on the Look Into Person dataset, which includes annotations for 2D human pose estimation and body part segmentation. Our framework achieves a mean accuracy of 74% and a mean intersection over union of 60%, outperforming state-of-the-art methods such as SCHP.

50 - Cutting-Planes for Arc Flow Formulations from Path Flow Formulations

Anthony Karahalios, Carnegie Mellon University, Pittsburgh, PA, United States, Willem-Jan van Hoeve

An arc flow formulation is an integer linear program that has a variable for each arc in a network. For directed acyclic networks, there exists an equivalent path flow formulation with a variable for each path through the network. We give an example of a cutting-plane for a path flow formulation for which there does not exist an equivalent cutting-plane in the arc flow formulation. We introduce a method that takes as input a cutting-plane in a path flow formulation and alters the network to permit an equivalent cutting-plane to be formulated in the arc flow formulation. We demonstrate the alteration method for Chvátal-Gomory rank-1 cuts derived from the path flow formulation. We give experimental results that show how the alteration method can translate such cutting-planes to improve the linear programming relaxation of an arc flow formulation for the capacitated vehicle routing problem.

51 - Food Detection from Continuous Glucose Monitoring Sensors Using Pretrained Transformer-Based Models

Abhimanyu Kumbara, WellDoc Inc., Columbia, MD, United States, Junjie Luo, Mansur Shomali, Anand Iyer, Gordon Gao

Healthy eating is a critical self-management behavior for a number of chronic health conditions such as diabetes and obesity. Manually logging meals is burdensome, and individuals with health conditions often do not consistently track their meals. Sporadic meal entries lead to sparse data, which may hinder data analysis, clinical decision support, and patient coaching. In this study, we present a novel approach for automatically detecting recent food intake solely based on continuous glucose monitoring (CGM) data.

We created three transformer encoder-based food event detection models: CGMEncoder, CGMEncoder+Time, and CGMEncoder+Time+Pretraining. The input to the models is a full 24-hour CGM trace, and the output is a discrete label indicating the hour (within a specified range) when a food event occurs. Our dataset consists of over 90,000 input-output pairs, partitioned into training, validation, and test sets. Each instance for the model includes CGM token sequences, hour-minute positional identifiers, and auxiliary time-related features capturing dietary event frequency and recency. The test accuracy increases from 63.0% with CGMEncoder model to 71.4% and 77.8% for the CGMEncoder+Time and CGM+Time+Pretraining models, respectively. These results demonstrate the importance of temporal encoding and pretraining for building robust and generalizable food event detection models using CGM data alone. These types of food event detection models can be embedded in digital health tools to automatically collect valuable food data, which can then be used for clinical applications.

52 - Decarbonizing Freight Transportation: Joint Optimization of Intermodal Cargo Routing, Scheduling, and Service Networks

Zeyu Liu, West Virginia University, Morgantown, WV, United States

Intermodal transportation, which exploits the strengths of alternative modes of transpiration, such as rail and water, offers a promising solution for the growing freight transportation sector to reduce carbon emissions and operating costs. Optimization of the intermodal transportation system has been widely studied in the literature, but conventional approaches typically focus on the broad averages of emission and cost statistics or rely on existing inflexible timetables that lead to suboptimal results. In this study, we optimize the U.S. intermodal system consisting of highway, railway, and waterway freight transportation. We construct a mixed integer programming model on a multi-layered transportation network to jointly optimize freight routing, scheduling, and service network design decisions. The model minimizes intermodal operating costs, carbon emission costs as carbon taxes, and late penalties for delayed freight arrival. The complexity of the intermodal system imposes great computational challenges on state-of-the-art solution methods. To address this, we investigate mathematical properties of the model to explore structural benefits. We further propose a vehicle scheduling decomposition approach that expands the Lagrangian cutting plane methods towards nonlinear linking constraints. Structural properties are integrated into the decomposition to derive novel algorithms. Real-world intermodal infrastructure, cost, and freight projection data are collected to validate the model and the algorithms through comprehensive experiments. Carbon emission data from combinations of vehicle engines and fuel types are utilized in large-scale case studies to provide insights on intermodal decarbonization, targeting a carbon-neutral U.S. freight transportation sector by 2050.

53 - Collaborative New Product Development: the Implications of Long-Term vs. Short-Term Contract Choice on Cost-Sharing in Vertical Channels

Abhishek Roy, Temple University Fox School of Business, Philadelphia, PA, United States, Ganesh Balasubramanian

Problem Definition: It is well known that a buyer expects price commitment from upstream supplier when they collaborate to improve the product quality. This insight is based on the notion that an upfront price commitment from the supplier allows the buyer to confidently co-invest without having to worry about the supplier exploiting the buyer with a future price rise. Existing research on supplier-buyer co-creation has not addressed the impact of the buyer's ability to influence the supplier's demand in multi-period selling environments by carrying strategic inventory. We investigate how the possibility of strategic inventory influences buyer's co-creation effort, product quality, and the supplier's contract choices between short-term transactional and long-term commitment contracts.

Methodology/Results: Using a stylized game-theoretic model, we consider the collaborative interactions between a supplier and a buyer over a two-period horizon in which demand is deterministic. Our analysis reveals that the buyer does not benefit from price commitment, and its collaboration effort decreases with price commitment for sufficiently low values of holding cost. In addition, we find that the supplier may prefer a long-term commitment contract despite losing the flexibility to vary prices.

Managerial Implications: Considering the impact of short-term dynamic contracts on supplier-buyer collaboration efforts in a multi-period setting, we contribute to the literature on collaborative product development and strategic contracting. Our research enables suppliers and buyers who engage in collaborative product development to optimally determine their effort and contract choices.

54 - Uncertainty-Aware Large Language Models for Risk-Sensitive Motion Planning in Autonomous Driving

Yuming Chang, Georgia Institute of Technology, Atlanta, GA, United States, Yongyang Liu, Ruining Yang, Nastaran Moradloo, Srinivas Peeta

Autonomous vehicles (AVs) operate in highly dynamic environments where accurately identifying and interpreting surrounding vehicles and human behavior is critical for their safety. Current models predominantly focus on spatial-temporal data, often neglecting nuanced contextual and interpretive reasoning required to handle uncertainties effectively. This research addresses critical gaps in quantifying environmental and behavioral uncertainties using Multimodal Large Language Models (MLLMs). We propose a multimodal contextual reasoning framework that uses MLLMs to detect, classify, and semantically interpret the behaviors and intentions of surrounding vehicles and humans by utilizing multimodal environmental data including camera, LiDAR, and radar. Our probabilistic prediction framework employs MLLM-driven reasoning to explicitly quantify uncertainties by providing predictive entropy and confidence scores, which are crucial for navigating uncertain and dynamic driving scenarios. **Further, Thenmore**, we develop a Bayesian decision-making system explicitly integrating these quantified uncertainties to enhance real-time, safety-critical decision-making. The effectiveness of this approach is validated through extensive simulation and evaluation, employing metrics such as time-to-collision and collision rates within simulated environments, demonstrating its usability and reliability in improving the safety of autonomous vehicles.

55 - Rethinking Risky Portfolio Allocation as An Online Mixed Integer Stochastic Bandit Problem

Jeffrey Varner, Cornell University, Ithaca, NY, United States

This paper presents a framework for online portfolio construction by redefining the mean-variance allocation problem as an online mixed-integer stochastic multi-armed bandit problem. Each arm represents a binary vector representing a candidate asset subset, with rewards computed via Cobb–Douglas utility maximization under budget constraints. Historical market data and a risk-aversion parameter inform the exponents in the utility function, which measure investor preference. We employed an ϵ -greedy algorithm with a decaying exploration rate and an analytical solution for the utility maximization problem, to compute the expected utility (reward) from each arm. This approach computes the optimal subset of assets to include in a portfolio and the number of shares of each asset to hold in the portfolio as a function of time.

We tested our approach on static and dynamic rebalancing problems with portfolios from S&P 500 constituents and ETFs during a 2024 out-of-sample period. Static analyses showed the bandit-driven maximum-utility portfolios approximate the efficient frontier, achieving over 90% similarity to the nearest mean-variance optimal weights while excluding non-preferred assets. By adjusting the risk-aversion parameter, we traverse the frontier: lower risk-aversion leads to risk-seeking allocations with higher expected returns, while higher risk-aversion favors portfolios composed of lower-risk assets. Finally, the algorithm identified optimal portfolios despite noisy market parameter estimates, demonstrating robustness against model uncertainty.

This bandit formulation offers an adaptive online learning approach to portfolio construction, integrating investor preferences for responsive portfolio management. Our findings highlight its potential for real-time personalization and suggest paths for enhanced context models and utility specifications.

56 - A Data-Driven Graph Model Approach to Critical Care Patient Clinical Trajectory Prediction

Honglin Du, University of Florida, Gainesville, FL, United States, Muxuan Liang, Xiang Zhong

The advent of increasingly available patient health data has created opportunities for critical care medicine to embrace modern data-driven decision support. However, existing methods, such as rule-based scoring systems or machine learning-based classification models, often focus on uni-dimensional patient outcomes such as mortality, lacking the granularity necessary to understand a patient's overall health during their hospital stay. This research focuses on predicting a patient's clinical trajectory, described by key health indicators readily accessible through electronic health records, by estimating correlations and causal relationships among these health indicators over time. To overcome the challenges associated with multi-dimensional, longitudinal, and heterogeneous patient data, we develop an analysis pipeline that first identifies distinct patient clusters with similar clinical trajectories through temporal pattern analysis. Then, for each cluster, we implement a structural equation model (SEM) to capture both immediate and time-lagged dependencies among variables, which offers a deeper understanding of the dynamic relationships. Directed acyclic graphs (DAGs) are used to visualize causal and correlative relationships, providing intuitive insights into how vital health indicators interact and form a patient's clinical trajectory. This integrated approach, which combines clustering, SEM, and DAG visualization, enables more refined modeling of patient clinical trajectories, yielding predictive insights tailored to specific patient groups. The approach has the potential to guide personalized interventions, ultimately improving patient outcomes in critical care settings.

57 - From Liquidity to Legacy: Designing Value-Centric Tokenomics for Sustainable Blockchain Innovation and RealFi

Ruizhe Jia, Stanford University, Palo Alto, CA, United States

The rapid liquidity available in cryptocurrency markets presents a double-edged sword: while facilitating access to capital, it often creates incentive structures that favour short-term gains over sustainable, long-term innovation. This "incentive-liquidity mismatch," particularly prevalent in current token financing models, can lead to premature exits by founders and investors, hindering the development of robust products—a critical challenge for ambitious goals like RealFi which require sustained effort and real-world integration.

This project seeks to address this fundamental problem. We argue for a **paradigm shift away from simplistic, time-based token vesting towards novel tokenomic models intrinsically linked to verifiable, real-world value creation, such as protocol revenue or user adoption metrics**. Our core objective is to design and analyze new crypto-native corporate finance and governance structures that effectively incentivize long-term commitment and reward the delivery of tangible, impactful products.

Employing a dual approach, we will first **empirically analyze** existing crypto projects to quantify the shortcomings of current models. Second, leveraging **mechanism design and game theory**, we will **develop and theoretically validate** these new value-centric tokenomic frameworks.

58 - Critical Configurations: Unraveling the Emergence of Outlier Technology Entrepreneurship in Crowdfunding

Xuanye Li, New Mexico Institute of Mining and Technology, Socorro, NM, United States, Christopher Crawford

Outlier technology entrepreneurship (TE) represents an extraordinary yet underexplored phenomenon within the broader study of superstar entrepreneurial success. This study examines the critical configurations of endowments and engagement that drive outlier outcomes in reward-based crowdfunding.

A robust dataset of 39,908 observations from 2014 to 2022, scraped from the population of Kickstarter.com campaigns, serves as the empirical foundation for this research. To uncover the generative mechanisms of outlier outcomes, we utilize three analytical methods, including fuzzy-set qualitative comparative analysis (fsQCA), Machine Learning—Word Embedding Models, and the analysis of heavy-tailed distributions.

Among the five significant findings, campaigns that receive a platform endorsement (i.e., identified by Kickstarter as 'Projects We Love') and post frequent interactions with backers via frequently asked questions (FAQs) emerged as key drivers of outlier success. We explicate the four other critical configurations that lead to outlier funding outcomes in the findings section.

Our findings reveal distinct configurations of factors for prioritizing resource accumulation that significantly enhance the likelihood of achieving exceptional funding success, offering actionable insights for entrepreneurs, and extending the theoretical understanding of outlier dynamics in entrepreneurial ecosystems.

59 - Adaptive Intelligence in Service Operations: Reinforcement Learning vs. Traditional Staffing Heuristics

Shenghan Xu, University of Idaho, Moscow, ID, United States

Effective service operation staffing is critical in managing fluctuating demand and ensuring high service quality, particularly in call centers. Traditional staffing strategies—**level, chase, and mixed approaches**—are widely adopted due to their simplicity and compatibility with spreadsheet-based planning tools. However, these rule-based heuristics may fail to adapt dynamically to real-time and nonlinear demand patterns, leading to suboptimal performance.

This study examines the application of **reinforcement learning (RL)** as a data-driven approach to staffing decision-making in service operations. Using two years of call center operational data, we evaluate the performance of RL-based staffing policies against conventional level, chase, and mix strategies. The RL model is trained to optimize staffing decisions by learning from historical call volume patterns and agent performance metrics, with the objective of minimizing cost while maintaining service level targets.

To ensure practical relevance, we develop an **Excel-based simulation framework** to emulate call arrival processes, agent availability, and customer service levels under each strategy. We further demonstrate how RL policies can be approximated and operationalized in Excel using rule extraction and scenario mapping, bridging the gap between advanced AI techniques and managerial implementation tools.

Results indicate that RL-based staffing consistently outperforms traditional heuristics across multiple demand scenarios, especially under high-variance or peak-load conditions. Moreover, we highlight a feasible pathway to integrate RL insights into existing Excel-based planning systems, providing a scalable solution for service operations that seek to leverage AI without overhauling their legacy infrastructure.

60 - Beyond Binary Classification: a Comprehensive Multi-Criteria Assessment System for Programming Assignments

Dan O'Leary, Auburn University, Auburn, AL, United States

The effective assessment of coding assignments presents several pedagogical and practical challenges. Traditional automated assessment systems reduce various educational objectives into a binary classification problem. The result fails to assess implementation quality and lacks actionable feedback, both essential to improving learning outcomes.

We present a hybrid system addressing the fundamental trade-off between assessment throughput and educational quality control. By integrating code execution tests with structured LLM analysis, it enables multi-criteria evaluation while improving operational efficiency. Declarative rubric statements designed for True/False evaluation focus on implementation logic and structural requirements, while correctness is verified through output comparisons and unit tests. A threshold-based scoring mechanism requires submissions meet minimum competence requirements to receive partial credit, enforcing learning outcomes with more nuance than any binary approach.

The framework automatically flags submissions for human review when confidence is low or anti-patterns are detected. This separation of concerns positions the LLM as a structured, rubric-aligned evaluation assistant rather than an open-ended judge, preserving human control over educational standards.

Empirical validation using nearly 1,000 programming submissions demonstrates significant improvements in assessment consistency, depth of feedback, and grading efficiency. Students receive detailed, actionable feedback covering both correctness and implementation improvements, even when code execution fails.

While developed for introductory Python courses, the framework is readily extensible within software engineering education by adjusting the number, granularity, and abstraction level of rubric items. Moreover, this approach shows broader promise for assessment applications that are similarly concrete and rule-governed, in domains where systematic criteria can be clearly defined.

61 - Linear Transformers as Vector Autoregressive Models

Jiecheng Lu, Georgia Institute of Technology, Atlanta, GA, United States, Shihao Yang

Autoregressive attention-based time series forecasting (TSF) has gained growing attention, with mechanisms such as linear attention sometimes outperforming standard attention. However, deeper Transformer architectures often conflict with autoregressive objectives, obscuring the vector autoregressive (VAR) structure inherent in linear attention and limiting their ability to model the underlying data-generating processes. In this work, we first demonstrate that a single linear attention layer can be interpreted as a dynamic VAR model. We then show that classic multi-layer Transformers introduce structural mismatches that reduce both interpretability and generalization. To overcome this, we propose a rearranged architecture, modifying the order of MLPs, attention modules, and data input, that allows multi-layer linear attention to align with a VAR structure. Building on this insight, we introduce Structural Aligned Mixture of VAR (SAMoVAR), a linear Transformer variant that incorporates interpretable, dynamic VAR weights for multivariate TSF. By structurally aligning the model with autoregressive objectives, SAMoVAR achieves superior performance, interpretability, and computational efficiency compared to state-of-the-art TSF methods.

62 - Disclosing AI, Questioning Authenticity: How Transparency Shapes the Value of AI-Generated Professional Work

Angel Hsing-Chi Hwang, University of Southern California, Los Angeles, CA, United States

This work presents two series of studies that examine how knowledge workers and their audiences perceive AI-generated content and how these perceptions shape professional outcomes. In the first series, we invited professional writers to use AI-assisted tools in their writing process and reflect on their experiences. Writers generally reported that their AI-assisted work still felt authentic. However, they expressed concern that clients or audiences might devalue their output upon learning that it contained AI-generated content. We analyzed data from 9,149 freelancers across two major online platforms to understand this tension. We found that disclosing AI use had divergent effects across domains: in creative fields, disclosure was associated with lower pay, while in non-creative fields, such disclosure correlated with higher compensation, suggesting that AI was seen as a value-add in most domains but a threat in creative ones. We also observed that freelancers who disclosed AI use were often novices with limited platform experience. To further investigate how disclosure interacts with the reputation of workers, we ran online experiments using synthetic freelancer profiles that varied by disclosure and popularity. Results showed that participants strongly preferred hiring popular freelancers regardless of AI disclosure; for less popular freelancers, disclosing AI use significantly reduced their likelihood of being hired. Together, these findings highlight the complex, context-dependent nature of AI disclosure and raise critical questions for designing transparent yet equitable systems for AI-mediated work.

63 - Do You Reason like Me? Evaluating Chain-of-Thought Alignment with Linguistic Personas in LLMs

Kezia Oketch, University of Notre Dame, South Bend, IN, United States

Chain-of-thought (CoT) prompting has emerged as a powerful method for eliciting structured reasoning from large language models (LLMs). Yet little is known about whether these reasoning paths align with users' sociolinguistic identities, particularly in multilingual, low-resource settings. This paper investigates whether persona-aware prompting can improve both the accuracy and epistemic quality of LLM-generated CoT explanations.

Using a Swahili-language dataset of 2,170 free-text health responses annotated with demographic and linguistic features, we define speaker personas based on *Sheng* (urban vernacular) usage, tribal lexicons, code-mixing, and loanwords. We prompt three instruction-tuned LLMs (GPT-4o, Llama 4-Scout, Qwen3-32B) to score each response and generate an accompanying explanation under five conditions: (1) no CoT (score-only), (2) standard CoT, (3) persona-tuned CoT, (4) misaligned CoT (incorrect persona cues), and (5) few-shot persona-primed CoT.

We evaluate performance using Δ Score (Human – LLM), and introduce a multi-layered analysis of CoT explanations, assessing semantic relevance, hallucination rates, tone appropriateness, and stylistic alignment with each persona. Our results reveal that persona-tuned prompting improves alignment for marginalized linguistic profiles and yields explanations that are more culturally respectful and contextually grounded. However, these benefits vary by model and task (trust in doctor, numeracy, subjective literacy, and anxiety visiting a doctor), and misaligned prompts can degrade reasoning quality or amplify stereotypes.

This study contributes a new fairness diagnostic at the reasoning level and offers persona-aware prompting as a lightweight, deployable strategy to enhance epistemic inclusion and cultural sensitivity in LLM explanations.

64 - Determining the Optimal Material Cost Threshold and Its Corresponding Time for Battery Plant Investment

K Jo Min, Iowa State University, Ames, IA, United States, Olivia Bujak, Eli Sallis, Gazi Nazia Nur, Mohammad Ahnaf Sadat, John Jackman

Due to substantial increases in electric vehicle (EV) demand, when to invest in battery plants for EV's has become an economically critical and much strategic business decision. A key reason that such decision-making is challenging is that the costs of the essential materials such as Lithium and Nickel are highly volatile over time. Under

these circumstances, we investigate how the optimal material cost threshold and its corresponding time for the battery plant investment can be determined based on a class of stochastic optimal controls with geometric Brownian motion processes representing the price evolution of Lithium or Nickel. By solving resulting second order stochastic differential equations, we demonstrate the optimal timing of such investment decisions and the value of the corresponding option to invest. Through a combination of empirical and simulated data, we provide numerical examples in the context of cost fluctuations of Lithium or Nickel. Economic implications and managerial insights are also presented.

65 - Analysis and Discussion of the Wga Strike of 2023

Lara Inci Kullu, Georgia Tech , Atlanta, GA, United States

In this paper, we study the Writers Guild of America strike of 2023. Writers are critical suppliers of the entertainment industry, which is a major service industry in the U.S. The strike affected not just the movie and television studios, but also the whole economies of California, Georgia, and New Mexico, placing the economic cost at more than 6 billion dollars. We look at the problem with analytical tools from labor economics and discuss equitable solutions for the industry as a whole and future research directions.

66 - Scale-Aware Mosaic Games for Multi-Level Strategic Risk and Resilience Design

Quanyan Zhu, New York University, New York, NY, United States

Evaluating strategic risks in large-scale complex systems poses significant challenges due to expansive attack surfaces, evolving threat models, and intricate interdependencies. Traditional risk assessment techniques often fail to scale or adapt effectively in the face of such complexity. Game theory offers a rigorous foundation for modeling adversarial behavior, enabling analysts to anticipate and mitigate threats through equilibrium concepts and incentive-aligned strategies. Yet, applying these tools directly to high-dimensional systems, such as cybersecurity infrastructures or interdependent critical infrastructures, frequently encounters computational and representational bottlenecks.

This talk introduces Mosaic Games, a multi-resolution dynamic game framework designed to overcome these limitations. Mosaic Games enable analysts to dynamically “zoom in” to examine local subsystems with fine-grained precision, or “zoom out” to evaluate global, systemic interactions. The framework supports hierarchical modeling, composability, and modular computation, facilitating scalable learning and real-time strategic adaptation. It provides a principled means to bridge local agent behavior and emergent global phenomena within a unified strategic architecture.

We illustrate this framework through its application to cybersecurity risk assessment, demonstrating how Mosaic Games facilitate both threat quantification and the design of resilient network architectures. Moreover, this approach generalizes to other domains characterized by nested or distributed decision-making, and it offers novel tools to capture systemic risks, where small perturbations yield cascading failures, and holonic risks arising from interactions among layered subsystems. By integrating granular and panoramic views of strategic behavior, Mosaic Games offer a compelling new lens for understanding and managing risk in today's adversarial, interconnected environments.

67 - Federated Short-Run Process Monitoring

Alina Gorbunova, Georgia Institute of Technology, Atlanta, GA, United States, Zihan Zhang, Kamran Paynabar

Short-run productions are common in manufacturing systems which require customization and flexibility under limited demand and resources. However, traditional statistical process control charts cannot handle such scenarios, since they either ask for data transformation or have strict assumptions which misalign with real-world applications. To address these limitations, we propose a federated framework to collect process parameters across multiple distributed product lines and create process monitoring baseline models for each line. Specifically, a Gaussian Mixture Model uses the gathered local information to estimate each line's parameters at the global server. By running the clustering algorithm, local sites receive the process parameters. Next, we personalize the local model based on the random effect model. Finally, we validate the model by using simulations and a case study.

68 - Graph Network-Based Transmission Line Screening for the Unit Commitment Problem

Farhan Hyder, Rochester Institute of Technology, Rochester, NY, United States

Solving the transmission-constrained unit commitment (TC-UC) problem efficiently for high quality solutions is one of the biggest challenges faced by independent system operators today. One way to achieve this is by removing superfluous network constraints from the problem. Several approaches have been used to identify such constraints. However, these methods are either too conservative or fail to maintain solution feasibility. In this work, a graph network-based approach is developed to identify the subset of transmission lines that can be safely removed from the TCUC problem. The idea is to capture the spatiotemporal relationship between past line loading levels, nodal demands, and future line loading levels to identify superfluous network constraints. To achieve this, a novel regression-based classification approach is developed, where the regression model is used to predict line loading levels, and different thresholds are applied to classify transmission line capacity constraints as necessary or not for the TC-UC problem. In addition, different classification thresholds are applied for peak and off-peak hours. The major advantage of this approach is that once trained, the model can be used under different classification thresholds. Numerical results on the RTS-96 and the Polish-2383 bus systems show that the proposed approach significantly improves computational efficiency without compromising the solution quality.

69 - Acquire, Build, OR Partner? An Agent-Based Framework to Facilitate Decision-Making on Companies Entry Strategies to Incorporate a Disruptive Technology in Current Operations

Yingzheng Liu, University of Virginia, Charlottesville, VA, United States, Zhen Liu, Shun Cao

Decision-making regarding the adoption of new technologies has become increasingly challenging. For example, a recent literature review finds that 70% of digital transformation projects ultimately fail to meet their goals, resulting in wasting trillions of dollars in global investment. Although firms must evaluate strategic options—internal development, acquisition, and partnership—when disruptive technologies arise, existing decision-support tools often overlook the dynamic interplay between legacy and emerging markets under conditions of high uncertainty and frequently require extensive data to yield actionable insights, limiting their real-world applicability. The model does not only capture interactions between a legacy market (the industry in which the firm is well-established) and a emerging market (the industry the firm is considering) by integrating firm-level attributes such as technological capability, adaptability, market-share dynamics, investment behavior etc., but also captures interactions including competition and cooperation within markets. Simulation results demonstrate the framework's effectiveness in replicating real-world firm behaviors and reveal behaviors such as how market volatility and resource endowments shape optimal entry strategies: high uncertainty conditions favors partnerships, whereas stable or low uncertainty conditions support internal development. Acquisition is also shown to be a versatile strategy across both high- and low-uncertainty environments. More interesting behaviors are to be explored since the model is designed to allow users to calibrate parameters to ensure replicability. This tool has the potential to empower decision-makers and policy-makers across different functional sectors to systematically anticipate market evolution and evaluate entry strategies, and ultimately make informed decisions with the help of enhanced strategic foresight.

70 - Modeling Port-Centric Supply Chain Disruptions: Lessons from the Francis Scott Key Bridge Collapse

Ziping Wang, Morgan State University, Baltimore, MD, United States, Di Yang, Qiannong Gu

The sudden collapse of the Francis Scott Key Bridge on the I-695 Baltimore Beltway in March 2024 presents an unprecedented disruption scenario, as it simultaneously severed a major regional transportation artery and obstructed maritime access to the Port of Baltimore, a vital international logistics hub. This event offers a rare and impactful opportunity to examine the cascading effects of critical infrastructure failure on multimodal supply chains.

This research investigates the collapse as a case study to model supply chain disruption dynamics at both the port operations level and across the broader logistics network. We develop analytical models to assess how logistics systems adapt in the face of compound transportation interruptions. Key objectives include predicting

response patterns, quantifying resilience gaps, and identifying mitigation strategies to enhance future disaster preparedness.

The findings of this study are intended to provide actionable insights for port authorities, logistics organizations, and policymakers in the development of more resilient and adaptable supply chain systems capable of withstanding rare yet high-impact disruptions. The research contributes to the growing discourse on infrastructure resilience and port-centric risk management.

71 - Supporting EV Tourism Trips Through Intermediate and Destination Charging: A Case Study of Lake Michigan Circuit

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

This research proposes a comprehensive framework for optimizing Electric Vehicle (EV) charging infrastructure along the Lake Michigan Circuit (LMC) to support ecotourism, incorporating both destination-based slow charging and corridor-based fast charging. The framework identifies the optimal locations and capacities of Level 2 chargers at key tourist destinations, as well as Direct Current Fast Chargers (DCFCs) along major travel corridors. Infrastructure planning is evaluated based on two primary objectives: (1) minimizing overall charging infrastructure costs, and (2) minimizing user inconvenience. It also identifies the most cost-effective allocation of a given budget between Level 2 and DCFC installations under different budget limits. By integrating the Level 2 and DCFC networks, the framework reallocates any unserved Level 2 demand at destinations to nearby DCFC stations while remaining within budget constraints. To address the optimization problem, the study employs Simulated Annealing (SA) and the Nondominated Sorting Genetic Algorithm II (NSGA-II), comparing their performance and solution efficiency. The case study focuses on two states along the LMC, Wisconsin and Michigan, and compares their results in terms of regional travel behavior and charging demand patterns. The results demonstrate that the proposed interconnected infrastructure meets EV demand across the LMC region under various budget scenarios. Overall, the findings highlight the importance of balancing investment between Level 2 and DCFC chargers to achieve optimal infrastructure outcomes. This balanced strategy supports EV adoption and ecotourism by ensuring reliable and accessible charging options, while also offering a foundation for future research on dynamic planning and broader regional applications.

72 - Integrating Structured and Unstructured Data to Estimate Baseline Status After Stroke: a Hybrid Machine Learning Model Enhanced by Language Models

Chung-Kai Chou, Hofstra University, Hempstead, NY, United States, Amir Gandomi, Bo Yang, Joseph Ohara, Jeffrey Katz, Jason Wang, Maria Sanmartin, Pina Sanelli

Accurately estimating a patient's baseline functional status at the time of stroke admission is critical for retrospective research and for risk adjustment in outcome prediction models. The modified Rankin Scale (mRS), a standardized measure of functional disability, is commonly used for this purpose. However, mRS scores are often missing or inconsistently documented in structured electronic health records (EHRs). This study presents a hybrid machine learning model designed to estimate baseline mRS scores by integrating structured EHR data with unstructured clinical text.

To address incomplete or missing data, we apply large language models (LLMs) to extract and impute mRS-related information from free-text clinical notes. Thus, LLMs are not used for prediction but serve as data augmentation tools, supporting missing data imputation and feature enrichment. This enables the creation of a high-quality dataset for supervised learning.

The final hybrid model combines structured variables with BERT-encoded clinical narratives and achieves significantly improved performance. Compared to structured-only models, the hybrid approach increases mRS classification accuracy from 80.2 percent to 94.5 percent using a random forest classifier.

This study shows how LLMs can enhance clinical datasets and support robust modeling of baseline functional status in stroke patients. The proposed framework offers a scalable solution for improving predictive performance in healthcare applications involving both structured and unstructured data.

73 - Contract-Based Multi-Agent Deep Reinforcement Learning for Integrated Timber Supply and Carbon Offset Management

Sorna Javadi, Department of Industrial and Systems Engineering, Mississippi State University, Starkville, MS, United States, Mohamad Dehghan Bonari, Bruno Kanieski Da Silva, Mohammad Marufuzzaman, Shaun Tanger

The coexistence of timber harvesting and forest-based carbon offset programs presents a growing challenge for managing supply chain decisions across multiple parties with conflicting objectives. Hence, this study investigates a competitive timber supply chain involving forest landowners, a wood-consuming mill, and a carbon aggregator. Modeled as a decentralized approach, each party in this supply chain aims to optimize its own objective without considering others' benefits. The landowners set stumpage price offers for their timber stands to maximize their land expectation values, while the mill plans to minimize its operating costs by purchasing from eligible landowners (i.e., only landowners with mature stands are selected). Meanwhile, the carbon aggregator offers carbon credit contracts to the landowners targeted by the mill, convincing them to defer harvesting to meet a carbon sequestration objective. Moreover, we extend this model by introducing a reservation-based option contract between the mill and landowners, modeled as a Stackelberg leader-follower game (mill as leader, landowners as followers). Under this contract, the mill pays an upfront premium for the right to buy timber at a predetermined future price, securing future supply. Landowners decide whether to accept the option by comparing the option-enhanced value of their forest (the immediate payment plus guaranteed future price) against potential future market returns or carbon payments. Both proposed models are solved and compared using a Multi-Agent Deep Reinforcement Learning (MADRL) framework, where each agent can learn a policy to optimize their long-term objectives.

74 - Dynamic Parking Pricing and Curb Space Allocation for Enhanced Urban Mobility Using Bayesian Optimization Approach

Prakash Ranjan, University of Connecticut, Storrs, CT, United States, Fatemeh Fakhroosavi, Krishna Murthy Gurumurthy, Abdelrahman Ismael, Nazmul Khan

Parking scarcity in downtown areas presents a major challenge for commuters seeking convenient spaces near their destinations. This study proposes a dynamic pricing optimization framework for garage parking in downtown Chicago, aiming to balance the often-conflicting objectives of parking authorities (i.e., maximizing revenue) and users (i.e., ensuring affordability especially relative to other choices). The framework incorporates a minimum occupancy constraint to promote efficient resource utilization while dynamically adjusting prices based on temporal demand variations. We employed simulated data from POLARIS activity-based transportation simulator, within a Bayesian optimization framework to determine optimal garage prices for both peak and off-peak periods. These optimized prices were then reintroduced into POLARIS to evaluate key performance metrics, including temporal and spatial demand distribution, occupancy rates, and revenue over a 24-hour period. The results show that reducing off-peak prices significantly increased demand by making garage parking more attractive and accessible, while higher peak-hour prices effectively aligned user demand with available capacity. The proposed approach improved peak-hour occupancy from approximately 70% to over 90% and increased overall utilization with a marginal reduction in revenue, demonstrating the effectiveness of the pricing strategy. Additionally, this study highlights the influence of allocating a portion of parking spaces to reserved users or specific vehicle types in enhancing curb space management and guiding travelers' parking choices. The findings support policy formulation aimed at reducing cruising behavior and alleviating traffic congestion, contributing to a more sustainable urban environment where parking is both accessible and affordable for a broader range of users.

75 - A Robust Approach to the Fleet Composition Problem with Truckload and Less-than-Truckload Shipping Options

Jiazheng Ling, The University of Arizona, Tucson, AZ, United States, Yue Wang

We investigate a two-stage robust optimization framework for the fleet composition problem under uncertain demand for multiple products, which integrates both dedicated truckload (TL) and supplemental less-than-truck (LTL) shipping options, with the objective of minimizing total costs. In the first stage, the model determines the composition of a dedicated fleet from a set of vehicle types with heterogeneous capacity, fixed costs, and dispatching costs, prior to any demand realization. In the second stage, once the demand is realized, the model optimizes dispatching strategies using a mix of TL and LTL services. To manage the complexity arising from the demand uncertainty, we approximate the problem by adopting its associated K-adaptability formulation, which partitions the uncertainty set into K subsets and identifies the best dispatching strategy for each subset. To solve the K-adaptability problem, we implement a scenario-based branch-and-bound algorithm that recursively solves a sequence of relaxed problems and identifies violating scenarios through a separation problem, branching when necessary to iteratively refine the solution. Our numerical

experiments demonstrate the effectiveness of this method in solving the fleet composition problem. Results from a comprehensive parametric analysis provide valuable managerial insights into the trade-offs among fleet investment, service flexibility, and robustness to demand variability.

76 - Waste-to-Energy Solutions in the Indonesian Energy System: Comparing Net-Zero CO₂ Alternatives in Electric Power Generation

Qoriatul Fitriyah, North Carolina State University, Raleigh, NC, United States, Matteo Nicoli, Ranji Ranjithan, Anderson de Queiroz

As one of the largest CO₂ emitters in the world (532 MtCO₂ in 2020), Indonesia ratified the Paris Agreement to reach zero CO₂ by 2050. In this work, we evaluate options to achieve zero CO₂ emissions in the country by 2050 in the power generation sector. We use Tools for Energy Modeling Optimization and Analysis (TEMOA) to explore four scenarios: Business as Usual (BaU), Net Zero (NZ), Net Zero with Waste-to-Energy (NZ+W) scenario, and Net Zero with WTE and Solar (NZ+W+S). The BaU scenario has the lowest cost (\$50.7 billion) but fails to eliminate CO₂ emissions at 451 Mt in 2050. The NZ scenario achieves zero CO₂ emissions by 2050 at a cost 13.4% higher than BaU. The NZ+W and NZ+W+S scenarios achieve zero CO₂ by 2050 at a higher cost (16.9% and 20.5% above NZ, respectively). However, NZ+W provides early emissions reductions and addresses Indonesia's 27.6 Mt annual waste problem by eliminating open pits, while NZ+W+S helps the government increase the renewable share. Our findings suggest that Indonesia can combine renewables and WTE to achieve zero CO₂ by utilizing combustible waste that would otherwise be disposed of in open pits.

77 - Short-Term Forecasting of Residential Energy Demand Using Machine Learning and Transformer-Based Deep Learning Models

Mohammad Hzayen, Binghamton University, Vestal, NY, United States, Yong Wang

The precise prediction of electricity demand is essential for maintaining stable power systems while enhancing their operational efficiency and sustainability. ARIMA and SARIMA models have gained popularity because they are straightforward to understand and require minimal computational power. Yet, they fail to effectively model the non-linear and seasonal patterns of real-world energy systems. Modern machine learning and deep learning techniques have developed stronger models that detect intricate time-based patterns while integrating weather data for prediction purposes.

This research compares various forecasting approaches, including Random Forest, XGBoost, and LightGBM from the ML family, as well as Long Short-Term Memory and Transformer architectures from DL. Real-world electricity load datasets are enriched by adding weather data variables and engineered temporal variables including lag terms, rolling averages, and time-based encodings.

The evaluation uses RMSE, MAE, MSE, MAPE, sMAPE, and R^2 as standard metrics for assessment. The results show that XGBoost and Transformer outperform other models in long-horizon and capturing high-variability.

This research develops a flexible and scalable forecasting framework incorporating weather data and time-aware modeling to enhance prediction accuracy. The study provides practical applications for grid planning, demand-side management, and renewable energy integration.

78 - Temporal Decomposition and Sensitivity-Driven Optimization of Stochastic Power System Capacity Expansion Using Distributed Computing

Keith Holliday, Michigan Technological University, Longmont, CO, United States, Ana Dyreson

Temporal Decomposition and Sensitivity-Driven Optimization of Stochastic Power System Capacity Expansion Using Distributed Computing

Solving high-resolution capacity expansion problems (CEPs) under deep uncertainty remains computationally intensive due to the scale and complexity of modeling long-term operations across multiple scenarios. This work proposes a scalable solution framework that combines scenario decomposition with temporal decomposition to improve tractability and parallel efficiency.

We begin with a sensitivity analysis to identify dominant sources of computational bottlenecks, revealing that storage dynamics, renewable portfolio standards (RPS) constraints, and inter-temporal linkages significantly impact solve time. Motivated by these findings, we extend the Progressive Hedging Algorithm (PHA) with a novel, cloud-based, distributed dual-axis decomposition strategy, where subproblems are partitioned by scenario and time blocks—enabling finer-grained parallelization.

Using a custom implementation built on Dask's distributed framework, we orchestrate concurrent solutions of scenario-time subproblems while preserving global feasibility through augmented coordination steps. Empirical results on the California Test System (CATS) test system and a modified CEP model demonstrate that this approach offers a favorable trade-off between reduced wall-clock time and minimal optimization loss compared to traditional scenario-only decomposition.

This work highlights the practical benefits of sensitivity-driven reformulation and distributed temporal decomposition in stochastic optimization for energy planning. Our findings support the broader adoption of dual decomposition approaches in high-performance cloud computing environments and lay the foundation for future extensions involving asynchronous coordination and adaptive decomposition schemes.

79 - Enhancing Last Mile Delivery Through Network-Aware Clustering

Wenda Zhang, ESRI, Redland, CA, United States, Mingdong Lyu, Heather Moe, WeeLiang Heng, Mingyi He

Last mile delivery is a critical and costly component of logistics, often hindered by urban congestion, geographic obstacles, and strict delivery schedules. Conventional clustering methods, such as k-means, rely on geometric proximity and neglect the complexities of real-world transportation networks, leading to inefficient routing. This study evaluates a network-aware clustering approach using the ArcGIS Pro Last Mile Delivery Solver, which incorporates actual road networks, travel times, and operational constraints. Comparative analyses demonstrate that network-based clustering significantly improves route compactness, reduces travel time and distance, and minimizes constraint violations. Additionally, it enhances route balance and fairness, addressing common issues such as fragmented service areas and overlooked geographic barriers. Through spatial analyses including Thiessen polygons, we show that clustering is not merely a preprocessing step but a strategic component that shapes route quality and operational performance. Our findings advocate for a shift from traditional geometric heuristics to GIS-informed clustering for more realistic and equitable logistics planning.

80 - Hallar: a New Low-Rank Solver for Solving Huge SDPs: Theory and Computation

Arnesh Sujanani, University of Waterloo, Atlanta, GA, United States, Renato D C Monteiro, Diego Cifuentes

This talk presents a new first-order method, HALLaR, for solving large-scale semidefinite programs (SDPs) with bounded domain. It is an inexact augmented Lagrangian (AL) method where the AL subproblems are solved by a hybrid low-rank method. In contrast to the classical low-rank method, the new method finds a near-optimal solution (with provable complexity bounds) of SDP instances satisfying strong duality.

The main computational workhouse of HALLaR is sparse matrix-vector multiplications, making it especially suitable for solving large-scale SDPs and GPU programming. Computational results comparing HALLaR to state-of-the-art CPU-based solvers on several large SDP instances show that the former finds higher accurate solutions in substantially less CPU time than the latter ones. For example, in less than 20 minutes, our method can solve, on the CPU of a personal laptop, a maximum stable set SDP with 1 million vertices and 10 million edges within $1e-5$ relative accuracy. Finally, we also briefly discuss that the GPU version of HALLaR is even faster than the CPU version, often achieving speedups of 50 to 100 times.

81 - Using Machine Learning Models to Understand the Factors Affecting Individuals Mode Choice for Health Care Visit

Md Sami Hasnine, Virginia Tech, Blacksburg, VA, United States, Tasnim Abbasi, Xiaomin Chen

Transportation systems, including healthcare facilities, are crucial in individuals' access to fundamental needs. Lack of transportation access is often a significant obstacle when visiting a doctor for scheduled appointments or when urgent care is needed. Public transportation accessibility is limited in many rural communities in the USA, where driving or ride-hailing is the only available option. There is a lack of studies which looked at how access to transportation choices impact individuals' access to healthcare facilities. To address this gap, this study developed various machine learning models (i.e., Random Forest, XGBoost, etc.) and AI models (i.e., Neural Network) to understand which factors impact the mode choice decision during their health care visits. This study uses the National Household Level Travel Survey (NHTS) 2022, the largest national household-level survey in the USA. Using this rich dataset, we particularly looked at individuals whose went for medical or dental care facilities and the modes for those visits. The model results show that individuals' access to cars, number of members in the household, income, home location, and a few land-use factors affect individuals' choice of mode while visiting medical or dental facilities. The model result suggests that subsidized transportation services should be available for low-income individuals to visit their healthcare facilities even if they do not have access to private cars. The findings have significant implications for making our transportation and healthcare systems more affordable and accessible.

82 - A Quantum Inspired Bilevel Optimization Framework for Bus Route Accessibility Improvement Using QAOA

Daniel Udekwe, Florida State University, Tallahassee, FL, United States

As quantum computing transitions from theory to application, its ability to address complex optimization challenges has attracted significant attention. In transportation planning, classical algorithms often struggle with the combinatorial complexity of large-scale problems, particularly in contexts like bilevel optimization, where strategic and operational decisions interact hierarchically. Hybrid quantum-classical algorithms, such as the Quantum Approximate Optimization Algorithm (QAOA), offer a promising alternative.

This study introduces a hybrid optimization framework that integrates QAOA into a bilevel model for bus route planning aimed at improving accessibility for seniors. The upper-level model (ULM), solved using a classical optimizer, minimizes average walking distances for residents in age-restricted communities. The lower-level model (LLM), responsible for optimizing route structure under practical constraints, is reformulated as a Quadratic Unconstrained Binary Optimization (QUBO) problem and solved using QAOA via IBM Qiskit.

Simulations on a grid-based scenario demonstrate that the quantum-enhanced approach can identify new stop placements that bring all target communities within a 400-meter walking radius, while keeping total route length within acceptable limits. This confirms the potential of QAOA to effectively handle combinatorial subproblems that are otherwise computationally intensive for classical methods.

By embedding quantum algorithms within a classical planning framework, this work advances scalable, equity-aware urban transit design. It showcases the feasibility of applying quantum techniques to real-world infrastructure challenges, offering a template for future integration of quantum computing into smart city planning.

83 - Nash Equilibrium of Codeshare Strategy and Flight Frequency Optimization in Airline Alliances

Ko-Wen Lin, National Tsing Hua University, Hsinchu, Taiwan, Yu-Ching Lee, Chun-Han Wang

This study develops a nonlinear optimization model to jointly determine codeshare cooperation levels and flight frequency decisions among airlines in an alliance. The model accounts for market demand, seat capacity, airport slot constraints, load factors, and revenue-sharing mechanisms between marketing and operating airlines. To capture strategic interactions in a competitive market, equilibrium outcomes are derived using Karush-Kuhn-Tucker (KKT) conditions within a Nash equilibrium framework. The objective is to maximize total alliance profit while balancing cooperation and competition among allied and non-allied carriers. Our results offer quantitative insights into partner selection, codeshare intensity, and frequency adjustments under shared operations. The model provides a theoretical foundation to support airline alliance decision-making and profit optimization, and can be extended in future work with numerical analysis or empirical validation.

84 - Network-Aware Contextual Bandits for Distributed Decision-Making

Chyun Deng, University of California, Berkeley, Berkeley, CA, United States, Huiwen Jia

We study dynamic decision-making in distributed environments and propose contextual bandit algorithms that leverage both inter-node similarities and heterogeneity to enable scalable learning. Motivated by multi-city pricing applications, we introduce two network-aware online learning algorithms: NetLinUCB, which maintains local parameter estimates and aggregates information via adaptive network weights, and Net-SGD-UCB, which performs local stochastic gradient updates and then constructs the confidence interval to perform uncertainty-aware exploration. Both methods decompose learning into global and local components. We establish theoretical regret bounds and show that our methods reduce the regret associated with the shared feature space from linear scaling in the number of cities $O(N)$ to sub-linear $O(\sqrt{N})$, significantly improving sample efficiency. We conduct extensive experiments and show that the proposed algorithms consistently outperform standard LinUCB, particularly in high-dimensional feature setting.

85 - Public Chargers for Electric Semis: a Bi-Level Optimization Approach

Mansimran Singh, University of Iowa, Iowa City, IA, United States

Electrifying long-haul freight hinges on deploying high-power chargers where fleets actually travel, yet most carriers cannot individually justify the capital or utilization risk of dedicated infrastructure. We model the required coordination as a bi-level principal-agent framework tailored to the I-10 corridor. In the upper level, a third-party developer (principal) selects candidate hub locations, sizes each site's charger array, and sets location-specific energy tariffs subject to contractual service guarantees on congestion and availability. In the lower level, multiple fleet operators (agents) solve continuous route-and-charge decisions that embed state-of-charge dynamics, heterogeneous values of time, and queueing-based waiting times. The principal's investment and pricing choices and the agents' operational responses are mutually dependent through congestion and energy demand, producing an optimization that links infrastructure decisions with day-to-day logistics behavior.

To navigate this coupling, we design an iterative bilevel decomposition: the principal's discrete-continuous planning problem is solved, holding congestion surrogates fixed; agents then independently update their optimal routing and charging strategies given the current hub design and prices; refined congestion estimates feed back to the principal stage, and the process repeats until decisions and congestion estimates stabilize. This approach preserves tractability while capturing the essential feedback loop among facility sizing, tariffs, and fleet routing.

The framework reveals how service-level guarantees, differentiated pricing, and modest charger utilizations can jointly align developer profitability with fleet welfare, thereby enabling coalition-based charging hubs without requiring each participant to bear full capital risk. Though calibrated to I-10, the model readily extends to broader corridor networks and evolving megawatt-scale technologies.

86 - Evaluating Discretization Effects in Bayesian Networks for Multi-Hazard Risk Assessment in Java, Indonesia

Himanshu Rana, NA, Kirkland, WA, United States, Rolf De By

Java, Indonesia is highly susceptible to multiple natural hazards, including earthquakes, volcanic eruptions, and tsunamis. These hazards often occur in interconnected sequences, forming disaster chains that significantly amplify environmental and societal impacts. Understanding the relationships among these hazards is essential for accurate risk assessment and mitigation. Hence, this study proposes a Bayesian Network (BN) based framework to model the probabilistic dependencies within these multi-hazard systems, with a focus on disaster chains involving volcanoes, earthquakes, and tsunamis. A key challenge in developing accurate BNs is the discretization of continuous geospatial and hazard data. While discretization is a common preprocessing step in BN modeling, it often leads to information loss, introducing bias into inference and decision-making. This study evaluates the influence of different discretization algorithms on model performance, which is frequently overlooked in environmental BN applications. The classification performance of the BN was assessed through Receiver Operating Characteristic (ROC) curve analysis, comparing discretization strategies to identify optimal methods for preserving data integrity. Additionally, spatial hazard mapping is conducted using ArcGIS to visualize the conditional hazard levels across the Java coastline. Results of the study demonstrate that the BN effectively captures the relationship of cascading hazards and that

discretization methods significantly influence predictive accuracy. The generated hazard maps reveal variable risk levels along the southern coastline, ranging from moderate to high. This study highlights the need for careful selection and documentation of discretization methods in BN-based hazard modeling to enhance reliability in disaster chain prediction and risk-informed decision-making.

87 - Automation and Robotics: the Future of the Warehouse

Cylia Blackmon, Kean University, Union, NJ, United States

Robotics and automation have revolutionized the way many supply chains and warehouses operate worldwide. They are increasing operational speed, decreasing order fulfillment time, and enhancing inventory management. This research explores the role of robotics and automation in warehousing and distribution, focusing on history, the variety of systems available, and associated advantages and disadvantages.

Automation dates back to the late 19th century with early innovations such as conveyor belts and barcode scanning, which optimized workflow. Modern advances in technology have companies considering automation's addition because of their ability to increase productivity as well as their high efficacy and efficiency. Modern systems such as autonomous guided vehicles (AGVs) and automated storage and retrieval systems (AS/RS) have been invented adopted. These findings suggest that by modernizing many of the repetitive and labor-intensive warehouse tasks, robots can assist in streamlining many processes, helping manage inventory, reducing associated costs, and delivering on consumer demand faster.

Even though automation has numerous advantages, including increased efficiency, higher profits, and continuous operation, it also has significant challenges. High initial costs, maintenance requirements, and the physical size of these systems often limit adoption, especially for smaller businesses. Additionally, the potential for job displacement concerns raises ethical issues about automation, prompting companies to consider retraining employees to include working with the new technology. Despite these challenges, automation is crucial in addressing warehouse labor shortages and meeting the increasing consumer demand. By understanding both the benefits and their associated challenges, organizations can make informed decisions about integrating robotics into their operations.

88 - The Unintended Consequence of Hawk-Eye Challenge

Shengshuo Xu, University of Science and Technology of China, Hefei, China, People's Republic of

The Hawk-Eye (HE) Challenge is widely adopted in sports when players dispute an umpire's call. Intuition suggests HE improves final call correctness by reviewing umpires' decisions, but this overlooks potential reputation loss for umpires when HE overrules their judgments. This study examines how HE influences umpires' strategic decisions, players' challenge behavior, and call correctness. Without HE, umpires prioritize maximizing correctness: those with high signal accuracy follow their received signals, while those with low accuracy rely on prior beliefs. With HE, umpires balance correctness gains against reputation loss from overruling. In this case, high-accuracy umpires still follow their signals despite possible HE overruling, but low-accuracy umpires may intentionally make decisions to trigger or avoid challenges, depending on reputation costs and players' challenge tendencies. Interestingly, we find HE can unintentionally lead relatively high-accuracy umpires to avoid challenges, undermining correctness compared to scenarios without HE. This work clarifies how interactions between umpire signal accuracy, reputation loss, and challenger signal accuracy affect game correctness—providing managerial insights for optimizing HE Challenge adoption in sports.

89 - A Stochastic and Equity-Aware Framework for Fare-Free Transit Zoning and Scheduling

Jiaqing Lu, Florida State University, Tallahassee, FL, United States, Qianwen Guo

This paper proposes the first analytical stochastic framework for optimizing the design and implementation of fare-free transit services. Focusing on a transportation corridor with two primary travel modes (auto and bus) the model divides the corridor into two sections: an inner zone offering fare-free transit and an outer zone maintaining a fare-based system. To capture the uncertainty in transit demand, a Geometric Brownian Motion process is used to represent the stochastic evolution of demand density, improving the model's realism. In the static model, we optimize the length and service frequency of the fare-free zone to maximize total social welfare. Results show that fare-free transit increases ridership, reduces automobile use, and enhances social equity. Joint consideration of social welfare and equity objectives leads to longer optimal fare-free zones than optimizing for social welfare alone. The dynamic model adopts a real options approach to address fare policy switching over time, offering closed-form solutions for optimal timing thresholds to activate or terminate fare-free service. These thresholds vary under social welfare-aware and equity-aware regimes and account for switching costs. Application to the Utah Transit Authority (UTA) Blue Line demonstrates that dynamic policies yield higher performance than static policies, with earlier implementation and longer fare-free periods under social welfare objectives.

90 - A Joint Design of Redundancy, Spares, and Trajectory for Permanent Lunar Habitats

Ahsanul Abedin, Texas State University, San Marcos, TX, United States, Tongdan Jin

Problem definition: This study addresses the challenge of designing sustainable lunar habitats by tackling two key questions: How can system reliability, spare parts, and spacecraft trajectory be co-optimized for deep-space habitation on the Moon? Is it feasible to minimize total launch mass while meeting stringent reliability and availability goals? **Practical relevance:** We present a novel Redundancy-Sparing-Trajectory Allocation (RSTA) model, integrating reliability optimization with space logistics theory. This integrated approach is crucial for space agencies and commercial entities aiming for cost-effective, sustainable lunar presence. **Methodology:** To address the combinatorial optimization, the RSTA model is solved in two phases: in Phase 1, allocating redundancy and spare parts for the lunar habitat's Environmental Control and Life Support System to minimize mass using two metaheuristic algorithms; in Phase 2, optimizing spacecraft trajectory, rocket staging, and propellant mass with Dijkstra's Algorithm to reduce launch mass in a cislunar network of 9 orbital nodes and 23 velocity changes.

Results: Experiments show strong interdependence between rocket staging, launch mass, mission duration, resupply lead time, and spacecraft trajectory. Longer missions favor spares inventory, and shorter durations prioritize component redundancy. A three-stage rocket design is optimal for balancing launch mass and cost effectiveness. High specific impulse propellants and optimized structural ratios significantly reduce launch mass. **Managerial implications:** The RSTA model represents an integrated approach for deep-space mission planning. It allows for building reliable, cost-effective lunar habitats by optimizing reliability-redundancy, spares stocking, and costly logistics. Future works can extend the model to interplanetary missions, including crewed flight to Mars.

91 - An Integrated Study of Data-Driven Thermal Modeling and Identification of Lithium-Ion Battery Packs

Saima Alam, University of Alabama in Huntsville, Huntsville, AL, United States, Rahul Bhadani

Accurate real-time temperature prediction is critical for ensuring the safety, longevity, and efficiency of lithium-ion battery (LIB), particularly under dynamic electric vehicle (EV) operations. Our work advances the electro-thermal modeling of LIBs by formulating a nonlinear state-space framework that captures cell-to-cell heat exchange, integrating a two-RC equivalent circuit model with polynomial approximations of heat generation based on state of charge (SOC). A comprehensive parameter identification pipeline is developed using experimental data from the open-source NASA battery dataset showing that unlike traditional approaches Neural ODEs not only predict the thermal behavior of LIBs from data, but also learn the underlying thermal dynamics, achieving sub-degree accuracy with RMSE as low as 0.007°C, and simultaneously identify parameters with high precision, reducing estimation error from 0.0052 to 0.0017 under constraints. Traditional approaches e.g. ordinary least squares (OLS) provides a fast, interpretable baseline through explicit weight coefficients, yet their linear nature cannot capture the battery's non-linear dynamics. Transformer models address part of this gap by modeling long-range temporal dependencies and lowering peak-temperature errors during high C-rate, but still treat the system as a black-box. In contrast, neuralODE offer high predictive power and dynamic interpretability, generating deeper insight into the thermal behavior and guaranteeing consistency with the governing physical laws. We demonstrate that neuralODE can uncover internal dynamics of the LIBS while maintaining a stable prediction of temperature over time. For application, these findings will be important moving forward for temperature management and tracking battery health through parameter estimation from the derived dynamic equations.

92 - Privacy-Preserving Automated Vulnerability Detection: Enhancing Organizational Security Through Federated Learning and Integrated Code Analysis

Ronia Arabian, University of South Florida, Tampa, FL, United States, Motahareh Pourbehzadi, Giti Javidi

Software vulnerability detection is crucial for safeguarding organizational assets and ensuring the reliability of complex systems. While recent advances in deep learning have shown promising results for automated vulnerability identification, many existing approaches overlook critical privacy and security concerns, particularly when leveraging data from multiple organizations. This research investigates privacy-preserving methods, specifically federated learning, that enable collaborative model training across organizations without sharing raw code data. By preserving data privacy, federated learning allows access to larger and more diverse datasets, addressing a significant barrier in vulnerability detection. Additionally, this research provides a comprehensive evaluation of various vulnerability detection techniques by analyzing source code, comments, and graph-based code representations to understand their individual and combined impacts on detection effectiveness. This study will also explore privacy challenges posed by potential adversarial attacks on federated frameworks and analyze defense mechanisms to enhance organizational software security. Through empirical evaluation on real-world datasets, this study aims to improve risk management strategies and advance privacy-aware vulnerability detection systems within enterprise environments.

93 - Appreciating the Complexity: A Systems Thinking Framework for Strategic Decision-Making and Performance Enhancement with Actionable Strategies that Align with the Organization's Operational Realities

Anwer Muhammad, Mount & peak media consultancy services LLC, Dubai, United Arab Emirates

The application of systems thinking in complex organizational environments, emphasizing the transition from soft systems methodologies to hard system modelling in real-world decision-making contexts. Soft systems thinking (SST), with its roots in interpretive paradigms, is crucial for navigating unstructured, human-centric problems involving multiple stakeholder perspectives and values. It enables appreciation, dialogue, and ideation in early problem definition stages. In contrast, hard systems thinking (HST) offers structured, quantitative approaches to model defined problems and optimize solutions through tools like simulation, system dynamics, and decision analytics. The integration of SST and HST fosters a bigger approach to change in a context of rich picture, where systems are not only modelled but deeply understood in their power and influence structures. Through this awareness, we used the methodology that supports iterative learning, adaptive change, and sustainable decision-making responses by breaking down actions into short- and long-term objectives. This concludes by examining how this continuum can guide enterprise transformation initiatives, enhance stakeholder engagement, and deliver Context-sensitive situation to solutions within complex, evolving ecosystems. Simulation modelling further enhances decision-making by providing dynamic foresight into complex systems, enabling organizations to align strategic objectives with operational realities.

References:

- [1] Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Irwin/McGraw-Hill.
- [2] Peter Check land, Sue Holwell. Information, systems and information systems: Making Sense of the field, International Journal of Information Technology Project Management, 13:1, (1-20), Online publication date: 01-Feb-1998

94 - Risk Disclosure Premium in the Cross-Section of Stock Returns: a Generative Topic Modeling Approach

Beichen Zhang, Stevens Institute of Technology, Hoboken, NJ, United States, Shanshan Yang, Steve Yang

We examine the pricing implications of financial risk disclosures by applying generative topic modeling to Section 1A of 10-K risk factor filings for S&P 1500 firms from 2010 to 2023. Using embeddings from FinBERT, GTE-Qwen, and SFR-Mistral, we quantify firm-level exposure to specific risk narratives and compute corresponding risk scores. These scores are used to construct narrative-based risk factors through a two-step empirical asset pricing framework. First, we test the return predictiveness of these factors using the Fama-French 5 factor model via long-short portfolio sorts. Second, we implement Fama-MacBeth cross-sectional regressions to evaluate their risk premia alongside the standard market, size, value, profitability, and investment factors. Our results reveal a robust risk premium in operational risk disclosure - which remains statistically significant even after applying the Bonferroni correction for multiple tests. In addition, several other disclosure factors, such as funding risk, intellectual property risk, and infrastructure risk, also show some level of significance. These findings suggest that firm-disclosed forward-looking risks, as identified through generative embedding model on firm risk disclosures, offer economically meaningful and statistically robust insights into cross-sectional return variation.

95 - Volatility-Aware Deep Recurrent Neural Network: Integrated Garch-Gru

Jingyi Wei, Stevens Institute of Technology, Hoboken, NJ, United States, Steve Yang, Zhenyu Cui

We introduce a novel Volatility-Aware Deep Recurrent Neural Network for financial volatility modeling and forecasting. The model integrates the Generalized Autoregressive Conditional Heteroskedasticity (GARCH(1,1)) formulation directly into the Gated Recurrent Unit (GRU) architecture. It yields a unified recurrent unit, integrated GARCH-GRU, that jointly captures both traditional econometric properties and complex temporal dynamics. This hybrid structure leverages the strengths of GARCH in modeling 'stylized facts' of volatility, such as clustering and persistence, while utilizing the GRU's capacity to learn nonlinear dependencies from sequential data. Empirical evaluations across multiple financial datasets demonstrate that the integrated GARCH-GRU model consistently outperforms classical GARCH-type models, standard neural networks, and alternative hybrid architectures in terms of MSE and MAE. Compared to a GARCH-LSTM counterpart, the proposed model achieves average MSE reductions of 21.55%, 11.91%, and 23.85% for 1-day, 3-day, and 7-day ahead forecasts, respectively. Notably, the GARCH-GRU exhibits faster adaptation to sharp changes in realized volatility, resulting in more accurate real-time adaptation to volatility shocks. As a practical application, we compute daily Value-at-Risk (VaR) forecasts for the S&P 500 index and observe lower violation ratios, further validating the model's effectiveness in dynamic financial risk management.

96 - A Hierarchical Bayesian Modeling Approach to Predict Upper Extremities' Fatigue Using Wearable Sensors in A Dynamic Order-Picking Task

Setareh Kazemi Kheiri, University of Pittsburgh, Pittsburgh, PA, United States

Predicting upper extremities' fatigue in dynamic tasks is crucial for human reliability analysis and system safety in warehouses and distribution systems. New nonintrusive technologies, such as wearable sensors, can facilitate worker safety monitoring. This study proposes a hierarchical Bayesian modeling approach that incorporates random slopes, intercepts, and nonlinear terms to enhance prediction accuracy for upper extremities' fatigue. Data were collected from a laboratory experiment involving 14 participants who performed order picking tasks over four days under varying load and pace conditions. Inertial Measurement Unit sensors attached to the wrist, trunk, and upper arm recorded accelerometer and gyroscope signals. The proposed hierarchical model utilized task conditions, individual characteristics, and signal features, with prior distributions of parameters continuously updated over time with new observations. The results demonstrate that incorporating random effects alongside fixed effects significantly improves the prediction accuracy. Updating prior distributions over time reduced prediction errors for subjective fatigue indicators, thus supporting proactive interventions. Our approach enables more precise and individualized fatigue prediction in dynamic tasks, contributing to improved system safety in smart warehouses.

97 - Optimizing Distributed Bio-Based Chemical Production with Stochastic Modeling and Scenario Analysis

Motahareh Kashanian, Iowa State University, Ames, IA, United States, Sarah Ryan

This poster presents a stochastic optimization framework for designing distributed manufacturing systems for biomass-based chemicals powered by electrochemistry and renewable energy. To address key uncertainties, such as renewable energy availability, process productivity, and the social cost of carbon, we use a two-stage stochastic programming model to support investment and operational decisions under diverse future conditions.

Scenario generation plays a central role. Renewable energy output is modeled using location-specific solar and wind data from NREL's System Advisor Model (SAM). Process productivity is treated as a stochastic variable reflecting technological uncertainties, while carbon pricing scenarios are drawn from recent climate-economic models. These scenarios inform second-stage decisions such as electricity sourcing and facility utilization.

The model identifies robust supply chain configurations that balance cost and emissions under uncertainty. Compared to deterministic models, the stochastic approach results in more flexible infrastructure decisions and highlights the trade-offs between up-front capital investment and long-term operational adaptability.

This work demonstrates how scenario-based planning and stochastic modeling can improve the resilience and sustainability of distributed chemical manufacturing systems.

Tuesday, October 28, 2:45 PM - 4:00 PM

TD01

Building A Level 3 A301

Cardinal Operations/LogixEDU

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Unlocking Full Optimization Potential with COPT**Tiancheng Zhang, Cardinal Operations, Shanghai, China, People's Republic of**

In this presentation, the COPT team will demonstrate new capabilities and functionalities introduced in the latest COPT release, including features like the nonlinear solver and GPU acceleration. Additionally, the speaker will share COPT's advanced features and performance through live notebook demos, showcasing real-world use cases such as drone trajectory optimization, portfolio optimization and beyond.

2 - LogixEdu: A Game-Based, Experiential Approach to Supply Chain Learning**Ziping Wang, LogixEDU, Baltimore, MD, United States**

LogixEdu transforms supply chain education through an interactive game that simulates the full journey from materials to manufacturing, distribution, and sales. By combining experiential learning with game-based design, it simplifies complex dynamics while fostering transferable skills such as adaptability, initiative, and teamwork. This innovation bridges classroom theory with industry practice, preparing students and professionals for real-world supply chain challenges in an engaging and accessible way.

TD04

Building A Level 3 A304

Data Science for Cyber-Physical Security and Resilience

Invited Session

Quality, Statistics and Reliability

Chair: Dan Li, University of Wisconsin-Madison, Madison, WI, United States

Co-Chair: Abhishek Hanchate, Texas A&M University, Bryan, TX, United States

Co-Chair: Navid Aftabi, University of Washington, Seattle, WA, United States

1 - Dynamic Watermarking for Cybersecurity of Closed- and Open-Architecture Machine Tool Controllers**Abhishek Hanchate, Texas A&M University, College Station, TX, United States, Satish Bukkapatnam**

Modern manufacturing systems are recognized as highly vulnerable to cybersecurity threats, with Machine Tool Controllers (MTCs) representing a key point of risk on the factory floor. Undetected cyber intrusions can jeopardize safety, disrupt productivity, and undermine system integrity. In response to these challenges, Dynamic Watermarking (DWM) has recently gained attention for its ability to provide provable security assurances across diverse domains such as energy systems and autonomous vehicles. This paper presents the first implementation of DWM tailored for manufacturing MTCs. The method is designed to detect cyberattacks in real-time, protecting MTCs from malicious interference while maintaining formal performance guarantees. Validation was conducted on both a proprietary (Siemens Sinumerik 828D emulator) and an open-source (Servo controller) control platform. Experimental results demonstrate that the DWM approach can reliably detect injection, replay, and flip attacks with an average latency of just 22 milliseconds, all while introducing less than 2 micrometers of deviation in the toolpath.

2 - Maintenance Modeling, Journey Control Charting, and Classification for Vulnerability and Blockchain Management**Theodore Allen, The Ohio State University, Main Campus, Columbus, OH, United States, Mohammad Abdullah, Vimal Buck, Zhizuo Chen**

We describe roles for security infrastructure relating to using a new type of "journey" control chart to monitor new cyber vulnerabilities and crumbs from blockchains. Also, we explore the roles of Markov Decision Processes in creating improved cyber-physical maintenance plans. Data from current systems is used to illustrate the methods.

3 - Stealth-Constrained Strategic Attack Generation in Integrated Cyber-Physical Systems via Reinforcement Learning**Navid Aftabi, University of Wisconsin-Madison, Madison, WI, United States**

Cyber-Physical Systems (CPS) are widely used in critical infrastructures and face cyberattacks that can inflict physical damage. In a CPS, attackers exploit vulnerabilities to compromise sensors and controllers, with the explicit goal of causing such damage. Consequently, novel mitigation strategies and anomaly-detection mechanisms must be developed under adversarial conditions. To expose weaknesses in current defenses, we propose a holistic attack-modeling framework that generates stealth-constrained adversarial policies in integrated CPS using reinforcement learning. Our framework serves as a tool for cyber-risk assessment of existing systems and for evaluating detection algorithms. It also enables automated, strategic adversary emulation in CPS, with direct applications to red-team testing and the design of more resilient detection mechanisms. We model the attacker's challenge as a Markov decision process (MDP) with stealth enforced via a cumulative detection-cost budget, incorporated into the reward through a Lagrangian relaxation. The resulting constrained MDP is solved by a Soft Actor-Critic algorithm augmented with dual ascent on the Lagrange multiplier, producing a policy that maximizes expected physical damage while minimizing detectability. Our results demonstrate that these strategic attacks can inflict severe damage on CPSs while remaining undetected, underscoring the urgent need for more advanced detection algorithms and further research in cyber-physical security.

TD05

Building A Level 3 A305

Service Science Award Session III

Award Session

Service Science

Co-Chair: Muge Yayla Kullu, University of Texas at San Antonio, San Antonio, TX, United States

Co-Chair: Weiwei Chen, Rutgers University, Piscataway, NJ, 08854, United States

1 - Dynamic Matching with Post-Allocation Service and Its Application to Refugee Resettlement**Soonbong Lee, Yale university, New Haven, CT, United States, Kirk Bansak, Vahideh Manshadi, Rad Niazadeh, Elisabeth Paulson**

Motivated by our collaboration with a major refugee resettlement agency in the U.S., we study a dynamic matching problem where each arrival (a refugee case) must be matched immediately and irrevocably to one of the static resources (a location with a fixed annual quota). In addition to consuming the static resource, each case requires post-allocation service from a server, such as a translator. Given the time-consuming nature of service, a server may not be available at a given time, thus referred to as a dynamic resource. Upon matching, the case waits to avail service in a first-come-first-serve manner. Bursty matching to a location may result in undesirable congestion at its corresponding server. Consequently, the planner (the agency) faces a dynamic matching problem with an objective that combines matching reward (pair-specific employment outcomes) with costs for congestion at dynamic resources and over-allocation of static ones. Motivated by observed fluctuations in refugee composition across years, we design algorithms without relying on distributional knowledge from past years' data. We develop learning-based algorithms that are asymptotically optimal in certain regimes, interpretable, and computationally fast. Our design is based on learning the dual variables of the underlying optimization problem; however,

the main challenge lies in the time-varying nature of dual variables associated with dynamic resources. To address this, our theoretical development draws on Lyapunov analysis, adversarial learning, and stochastic optimization. When tested on real data from the partner agency, our method outperforms existing ones, making it a viable candidate for replacing current practices upon experimentation.

2 - The Power of Static Pricing for Reusable Resources

Jiaqi Shi, Columbia University, New York, NY, United States, Adam Elmachtoub

We consider the problem of pricing a reusable resource service system. Potential customers arrive according to a Poisson process and purchase the service if their valuation exceeds the current price. If no units are available, customers immediately leave without service. Serving a customer corresponds to using one unit of the reusable resource, where the service time has a general distribution. The objective is to maximize the steady-state revenue rate. This system is equivalent to the classical Erlang loss model with price-sensitive customers, which has applications in vehicle sharing, cloud computing, and spare parts management. With general service times, the optimal pricing policy depends not only on the number of customers currently in the system but also on how long each unavailable unit has been in use. We prove several results that show a simple static policy is universally near-optimal for any service rate distribution, arrival rate, and number of units in the system. When there are multiple classes of customers, we prove that static pricing guarantees 78.9% of the revenue of the optimal dynamic policy, achieving the same guarantee known for a single class of customers with exponential service times. When there is one class of customers who have a monotone hazard rate valuation distribution, we prove that a static pricing policy guarantees 90.4% of the revenue from the optimal inventory-based policy. Finally, we prove that the optimal static policy can be easily computed, resulting in the first polynomial-time approximation algorithm for the multi-class problem.

3 - Data-Driven Stochastic Modeling Using Autoregressive Sequence Models: Translating Event Tables to Queueing Dynamics

Daksh Mittal, Columbia University, New York City, NY, United States, Shunri Zheng, Jing Dong, Hongseok Namkoong

While queueing network models are powerful tools for analyzing service systems, they traditionally require substantial human effort and domain expertise to construct. To make this modeling approach more scalable and accessible, we propose a data-driven framework for queueing network modeling and simulation based on autoregressive sequence models trained on event-stream data. Instead of explicitly specifying arrival processes, service mechanisms, or routing logic, our approach learns the conditional distributions of event types and event times, recasting the modeling task as a problem of sequence distribution learning. We show that Transformer-style architectures can effectively parameterize these distributions, enabling automated construction of high-fidelity simulators. As a proof of concept, we validate our framework on event tables generated from diverse queueing networks, showcasing its utility in simulation, uncertainty quantification, and counterfactual evaluation. Leveraging advances in artificial intelligence and the growing availability of data, our framework takes a step toward more automated, data-driven modeling pipelines to support broader adoption of queueing network models across service domains.

4 - Equity vs. Efficiency in Service Systems with Differentiated Demand

Hamid Arzani, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Ming Hu, Hossein Abouee-Mehrizi

In service systems, the pursuit of efficiency by prioritizing certain groups often raises concerns about equity toward those who are being prioritized. We study the efficiency-equity trade-off for a queueing system with two customer classes, distinguished by their arrival rate, urgency level (measured by per-unit-time delay cost), and initial well-being at the time of arrival to the system. We examine two types of policies among all work-conserving and non-preemptive priority policies: the efficient policy, which maximizes the system's overall expected utility by prioritizing the urgent customers, and the equitable policy, which maximizes the minimum expected utility of individual customers. We fully characterize the equitable policy and show that it agrees with the efficient policy when the system utilization is lower than a threshold and disagrees otherwise. We analyze the relative efficiency loss associated with the equitable policy, namely the price of equity (PEQ), and the relative equity loss associated with the efficient policy, namely the price of efficiency (PEF). We show that a slight increase in service capacity can lead to significant improvements in both the PEQ and PEF. We further show that the PEQ is significant when there is a moderate gap in delay cost rates and neither arrival rate significantly outweighs the other. In contrast, the PEF is significant when the delay cost rates are close, and the arrival rate of the urgent class significantly dominates that of the nonurgent class. Our results challenge the perception that efficiency and equity lie at the opposite ends of the spectrum.

TD06

Building A Level 3 A306

Volatility Modeling and Financial Risk Forecasting

Contributed Session

Contributed

Chair: Hyeontae Jo, POSTECH, Pohang, N/A, Korea, Republic of

1 - Mean Vector Shrinkage Estimation

Vali Asimit, Bayes Business School, London, United Kingdom, Ziwei Chen, Nathan Lassance

Shrinkage is a statistical method that was propelled by the *Stein's Paradox*, which showed that the standard *Maximum Likelihood Estimator* is not the estimator with the lowest estimation error. Shrinkage estimators have been used with great success in multivariate estimation such as multiple linear regression, mean vector and covariance matrix estimation, portfolio weights etc. The shrinkage tool combines two estimators, the most common choice and a low-variance target estimator, so that the combined estimator is more stable than the common/natural estimator. We provide four distribution-free shrinkage estimators for mean vectors and show their consistency properties under the most general setting where both the sample size and vector dimension get large. Such a setting is of great importance in practice, for example in portfolio theory, when the sample size is often smaller than the number of assets. We illustrate that such theoretical guarantees are reflected in synthetic and real data experiments, but we also show that our novel solutions outperform all existing methodologies. We choose some portfolio theory applications to explain the advantages of our estimators.

2 - Systematic Liquidity Risk Management: A Novel Perspective on Derivatives

Joseph Yi, Mizuho Financial Group, New York, NY, United States, Baeho Kim

Systematic liquidity risk poses a critical challenge in financial risk management, particularly during periods of market stress when liquidity suddenly dries up, as experienced during the Great Financial Recession. Liquidity shocks can propagate rapidly across asset classes, including derivatives, and exhibit persistence that undermines traditional risk models. This paper presents a novel risk management framework that bridges the gap between theory and practice by incorporating aggregate liquidity risk into derivatives pricing and risk assessment. Our approach integrates both physical and risk-neutral asset return dynamics to enable consistent risk pricing and measurement. We model systematic liquidity risk as a linear combination of two mean-reverting stochastic processes: a diffusive component and a jump component, using a double-exponential jump diffusion model. To demonstrate the tractability and applicability of our model in equity market, we propose a two-step calibration procedure for empirically capturing volatility skew. The first step estimates 'common' liquidity risk parameters using S&P 500 option prices, while the second step calibrates 'key' asset-specific parameters. We apply our framework to price a path-dependent contingent claim hedged with vanilla options and simulate idiosyncratic asset returns to generate a stressed P&L distribution for Value-at-Risk (VaR) estimation. Our findings validate the model's effectiveness, highlighting how the separation of common and idiosyncratic parameters enhances interpretability and practical implementation. This research offers both a theoretical foundation and a practical toolset for incorporating systematic liquidity risk into derivatives risk management policy to set risk limits, providing clearer insights into stress testing, pricing, and hedging in potentially illiquid market environments.

3 - High-Dimensional Portfolio Optimization Under Heavy-Tailed-Ness

Amirmohammad Ebrahimi, University of Florida, Gainesville, FL, United States, Hamed Amini, Hongcheng Liu

Portfolio selection under realistic market conditions often involves complex, non-smooth objectives and high-dimensional, heavy-tailed return distributions. We study the sample complexity of portfolio selection problems formulated via the Sample Average Approximation (SAA) method, including a regularized variation. We consider general convex composite objectives (possibly non-smooth) and derive non-asymptotic high-probability bounds on the optimality gap between the SAA solution and the true population-optimal solution. Our analysis uses a decomposition framework that yields bounds that are dimension-free with respect to the feasible region. The results

provide new theoretical insight into the reliability of data-driven portfolio construction under limited sample settings and offer practical guidance for model selection in high-dimensional, heavy-tailed environments.

4 - Advancing Yield Curve Forecasting: Deep Learning Nelson-Siegel Models and Macroeconomic Insights

Hyeontae Jo, POSTECH, Pohang-si, Korea, Republic of, youngju Ahn, MYUNGJUN KIM, Bong-Gyu Jang

This study empirically examines whether incorporating macroeconomic information into a Deep Learning Nelson–Siegel (DLNS) framework improves yield-curve forecasts, using monthly U.S. Treasury data. The benchmark DLNS model and its macro augmented variant, DLNS-X, both outperform the traditional dynamic Nelson–Siegel (DNS) model across all forecast horizons. Integrating CPI inflation and M2 money supply into the DLNS-X model further enhances medium-term and long-term predictive accuracy during periods of elevated short-term interest rates, indicating that these variables convey incremental information beyond that contained in the yield curve. In contrast, the performance gap narrows in certain economic regimes, lending support to the macro-spanning hypothesis that key macroeconomic news is already embedded in the curve. These findings confirm that deep-learning term-structure models can capture complex nonlinear dynamics while preserving economic interpretability, identify the conditions under which macro factors most improve forecast performance, and establish a foundation for the development of more refined forecasting methodologies.

TD07

Building A Level 3 A307

Digital Market Design and Emerging Commerce Models

Contributed Session

Contributed

Chair: Xin Liu, Renmin University of China, Beijing, N/A

1 - Platform Economics and Piracy Control: when Supply-Side Enforcement Backfires

Guangjun Huang, Central South University, Changsha, China, China, People's Republic of, Fuqiang Wang, Yunchuan Liu

The rise of digital platforms has fundamentally changed content distribution in the digital economy. In contrast to previous findings favoring supply-side measures, we discover that supply-side enforcement unintentionally reduces discovery system investment and weakens network externalities, diminishing product quality. Our results suggest a reconsideration of anti-piracy strategies in platform-dominated markets.

2 - Does Smart Contract Add Value to a Retailer with Buyback Risk?

Ao Zhuo, Renmin University of China, Beijing, China, People's Republic of, Wanshan Zhu, Zhengping Wu

Retailers face buyback risk in supply chains when manufacturers fail to honor their buyback commitments, as documented in litigation records. A smart contract can mitigate this risk by functioning as escrow to reserve a portion of the manufacturer's wholesale revenue in a dedicated account exclusively for buyback purposes. We focus on this escrow role of the smart contract to examine the value it brings to the retailer under both exogenous and endogenous buyback risk. Intuitively, the smart contract benefits the retailer by directly reducing buyback risk and hurts the manufacturer by posing extra financial constraints. Under exogenous buyback risk, which arises from uncontrollable external events, our analysis shows a counterintuitive result that the smart contract adds value to the retailer only if she is risk-averse, while always benefiting the manufacturer. In risk-neutral cases, the retailer suffers from the implementation of a smart contract because it allows the manufacturer to charge a higher wholesale price to extract profit. If risk aversion is above a threshold, a smart contract can create significant value for both the retailer and the manufacturer, leading to a "win-win" outcome. Under endogenous buyback risk, which stems from the manufacturer's financialization, the smart contract may lead to a "lose-lose" situation when an attractive investment opportunity exists, as the smart contract prevents the manufacturer from pursuing the more lucrative opportunity. Our research highlights the significant potential value of smart contracts, while advising supply chain managers how to avoid its pitfalls.

3 - Decentralized Platforms: Governance, Design, and Information Asymmetry

Wenxin Zhang, Syracuse University, Syracuse, NY, United States

In the decentralized platform economy, platforms often support two primary modes for distributing service providers' offerings to end users: service fee channel and reselling channel. This paper investigates the implications of platform channel design when a service provider possesses service quality information unknown to the decentralized platform. We analyze a signaling game between a service provider and a decentralized platform, in which the better-informed service provider first sets the service price, which it may use to signal private quality information. Based on the service price, the platform decides what units of service to deliver through the reselling channel. Finally, if the platform opens its service fee channel, the service provider decides what units of service to deliver through the service fee channel. We find that the effects of platform channel design on the service provider and the platform depend on three model parameters: the commission rate, the degree of channel substitution, and the degree of quality differentiation. Interestingly, using smart contract may narrow the region of the win-win outcome in which dual channel mode benefits both platform and service provider. Moreover, the platform cannot always induce the service provider to share quality information through smart contract; similarly, dual channel mode may even decrease the value of using smart contract for the decentralized platform.

4 - Smarter Games: AI Adoption in Video Game Development

Linghui Feng, Boston University, Boston, MA, United States, Qi Wang, Marcus Bellamy

We investigate how video game developers adopt and disclose the use of artificial intelligence (AI) in their production processes. Leveraging a large-scale dataset of games, developers, and publishers, we examine which organizational and collaborative characteristics influence developers' likelihood of disclosing AI use and how AI tools are applied across different aspects of game creation. Our findings highlight the determinants and visibility of AI adoption in creative industries, offering insights into how transparency and technological capabilities shape innovation in digital content development.

5 - Information Sharing and Channel Structures In Live Streaming Commerce

Xin Liu, Renmin University of China, Beijing, China, People's Republic of, Jianghua Wu

In recent times, online platforms have established live streaming channels through which manufacturers can advertise their products by paying the platforms service fee for live streaming. Manufacturers with direct channels consider utilizing live streaming channels through agency selling or reselling formats. The live streaming channel offers a comprehensive product introduction and an immediate purchasing experience for consumers, which can mitigate product uncertainty and enhance demand on the platform. As a consequence of the expansion of the brand, live streaming has the spillover effect of increasing demand in the direct channel. A game-theoretic model is constructed for the purpose of examining equilibrium pricing and live streaming efforts, as well as studying the interaction between selling format and information sharing. The results demonstrate that information sharing confers benefits to both the manufacturer and the platform in the context of agency selling. Conversely, under reselling, it is always beneficial to the manufacturer but harmful to the platform when the cost coefficient of live streaming is high. Furthermore, the manufacturer is inclined to pursue agency selling with a low commission rate. When the service fee increases, the manufacturer accepts a lower commission rate for agency selling. Furthermore, the platform has the ability to exert influence on the manufacturer to select the platform's preferred selling format by modifying the service fee. The findings of this study offer valuable insights to managers seeking to identify optimal strategies for selling formats and information sharing, with consideration of live streaming.

TD08

Building A Level 3 A310

Online Platforms and AI

Invited Session

Information Systems

Chair: Ramah Al Balawi, Baruch College, New York, NY, United States

1 - From Thrill to Theft: the Unintended Consequences of Online sports Betting**Ecem Basak, Baruch College, New York, NY, United States, SHAGUN TRIPATHI, Yi-Jen (Ian) Ho**

In 2018, in a landmark legislation, the U.S. Supreme Court ended a federal ban on sports betting, allowing individual states to regulate the activity. This prompted rapid adoption and fueled a surge in online, platform-mediated gambling. However, accessibility provided by online betting platforms raises growing concerns about broader societal impact. In this short paper, we examine whether the legalization of sports betting is associated with increases in crime—specifically burglary and robbery, offenses closely tied to financial strain and opportunity. Using a panel dataset covering U.S. states and the District of Columbia from 2018 to 2023, we employ staggered difference-in-differences models to estimate causal effects. We find that legalization is associated with statistically significant increases in both crime categories. Theoretically, we extend Routine Activity Theory and Strain Theory to digitally mediated environments. Our findings surface overlooked societal costs of legalization and platformization of gambling, and the expansion of digital markets.

2 - Multi-Agent Framework for AI Debiasing Human Creation in Commercial Practices**Chen Liang, The University of Illinois at Chicago, Chicago, IL, United States**

Concerns over fairness and accountability in content creation have surged with the rapid integration of artificial intelligence and human activities in the commercial usage scenario. Our study aims to design a multi-agent framework to debias human narratives using AI implementation, incorporated human-in-the-loop-evaluation, for detecting and mitigating bias in human-created commercial content. We also seek to investigate the alignment between AI models and human examinations to address if the perception gap exists between two subjects and thus broaden the understanding of human-AI hybrid creation and create a more feasible debiasing guidelines toward such collaboration.

3 - The Role of AI-Generated Links in Reshaping Collaboration Networks: Evidence from An Oss Platform**Zifeng Cheng, University of Illinois Chicago, Chicago, IL, United States, Yuheng Hu**

Open-source software (OSS) development plays a pivotal role in driving digital innovation, yet its progress is increasingly hindered by community fragmentation and weak cross-project connectivity. Emerging technologies such as Generative AI offer new opportunities to reshape developer interactions through their autonomy and creative capabilities. Among these, AI-generated conversational links—such as ChatGPT links—may enable more connected and decentralized collaboration by fostering transparent, interpretable, and participatory knowledge sharing. Unlike static documentation, these links dynamically externalize users' cognitive processes and illuminate the rationale behind decisions. We conceptualize these links as a novel form of social learning that reduces information asymmetries and promotes broader knowledge diffusion across developer communities. Empirically, we find that the use of AI-generated links reduces collaborator deep engagement and contributes to a more decentralized network structure. Moreover, our results suggest that users who adopt AI-generated links tend to occupy stronger bridging positions within the network, which serves as a driving factor in reshaping network structure. Overall, our findings contribute to the emerging literature on AI-generated content by shifting the focus from individual performance to relational and structural outcomes. At the same time, we introduce a network-based perspective on OSS innovation and highlight the potential of AI-enabled tools to reduce fragmentation and strengthen community connectivity.

4 - The Impact of Generative AI on Collaborative Open-Source Software Development: Evidence from GitHub Copilot**Fangchen Song, University of Texas at Austin, Austin, TX, United States, Ashish Agarwal, Wen Wen**

Generative artificial intelligence (AI) enables automated content production, including coding in software development, which can significantly influence developer participation and performance. To explore the impact on collaborative contributions and coordination dynamics, we investigate the role of GitHub Copilot, a generative AI pair programmer, in open-source software development where multiple distributed developers voluntarily collaborate. Using GitHub's proprietary Copilot usage data, combined with data from open-source repositories and a generalized synthetic control method, we find that Copilot use increases project-level code contributions by 5.9%. This gain is driven by a 2.1% increase in individual code contributions and a 3.4% rise in developer participation. However, these benefits come at a cost as coordination delays increase by 8% due to higher communication demands. Specifically, Copilot use is associated with more discussion volume, broader participation in discussions, and greater discussion intensity. This reveals a trade-off. While AI expands who can contribute and how much they contribute, it simultaneously slows coordination in collective development efforts. Despite this tension, the combined effect of these two competing forces remains positive, indicating a net benefit from using AI pair programmers. Interestingly, we also find the effects differ across developer roles. Peripheral developers show smaller gains in project-level code contributions but face greater coordination delays than core developers, likely due to limited project familiarity. In summary, our study underscores the dual role of AI pair programmers in augmenting contribution volumes and reshaping coordination dynamics in open-source communities and suggests potential implications for both practice and the broader open-source ecosystem.

TD09

Building A Level 3 A311

AI Transforming Platforms and Society

Invited Session

Information Systems

Chair: Zhechao Yang, George Mason University, Fairfax, VA, United States

1 - Labeling AI-Generated Content: Implications for Platforms, Creators, and Consumers**Yan Zhu, The University of Texas at Dallas, Richardson, TX, United States, Jianqing Chen, Srinivasan Raghunathan**

With the rapid evolution of generative AI, the line between human generated content and AI generated content has become increasingly blurred. To ensure transparency and trust, some social-media platforms, such as Facebook and Instagram, have introduced AI labels to inform users. However, the impact of such labels on different creators' creation effort remains unknown. At the same time, social-media content has become a valuable source of training AI models. Platforms have either directly monetized the content on the platform or used it to train their own AI models, creating a new revenue source besides advertising. This adds complexity to a platform's profit maximization problem, as it must balance user engagement and content quality. In this project, we study the impact of AI labels on platforms, content creators, and consumers using a stylized game-theoretic model.

2 - Pay Transparency and Workforce Dynamics: Evidence from It Labor Market**Haofei Qin, Georgia Institute of Technology, Atlanta, GA, United States, Jiayu Yao, Mingfeng Lin**

While proponents of pay transparency policies argue that disclosing salary ranges can reduce information asymmetry and promote equity in labor markets, the broader implications for employer behavior and workforce dynamics remain understudied. This paper explores how mandatory wage disclosure influences hiring and retention in the IT labor market by exploiting the implementation of a regional pay transparency law. We show how wage disclosure mandates reshape recruitment strategies and employee retention in high-skill labor markets.

3 - (How) Can LLMs Enhance Privacy Research?**Xiang Cheng, University of Maryland College Park, College Park, MD, United States, Wen Wang**

Understanding consumer privacy preferences and decision-making is central to how businesses and policymakers approach data governance, personalized services, and regulatory compliance. This study explores the potential of Large Language Models (LLMs) as scalable substitutes for human subjects in privacy research. We first assess whether LLMs exhibit a *privacy mindset*—human-like privacy preferences and decision-making—by benchmarking their responses against those of human subjects. Our findings reveal that LLMs not only demonstrate clear privacy preferences but also make reasonable, context-aware decisions. Interestingly, LLMs show higher privacy concerns than humans yet are more willing to share information in real-world scenarios. They also reflect both the *privacy calculus* (balancing risks and benefits) and the

privacy paradox (expressing concern while compromising privacy). These results suggest that LLMs possess a coherent privacy mindset, supporting their use as foundational tools for simulating human subjects in privacy research. To enhance alignment with human privacy behavior, we introduce a methodology that combines multi-persona modeling, Generative Adversarial Networks, and interpretable machine learning. This enables the generation of synthetic human subjects that closely mirror real-world privacy preferences. Extensive experiments show that our method significantly improves alignment between LLM and human responses across datasets and decision contexts. By establishing LLMs as viable proxies for privacy research, our study offers a scalable, cost-effective methodology for studying privacy behavior, with broad implications for academia, industry, and policy.

4 - Is Consumer AI Really Beneficial to a Firm?

Yue Li, Tianjin University, Tianjin, China, People's Republic of, Jianqing Chen, Srinivasan Raghunathan, Yan Zhu

Consumer AI, which is becoming increasingly prevalent in practice, significantly alters consumer purchasing behavior. A question that warrants further investigation is whether the adoption of consumer AI truly benefits the firm. We examine the optimal strategy for firms adopting consumer AI in a market where a manufacturer sells a product to consumers. We find that the manufacturer should adopt consumer AI (i.e., consumer AI is beneficial to the manufacturer) if the product base value is moderate, and otherwise, not adopting AI is the optimal choice (i.e., consumer AI hurts the manufacturer). Our findings provide guidance for the adoption of consumer AI by firms.

5 - Backfire of AI Marketplace Entry? Empirical Evidence from Start-Ups' Gpts Adoption

Xinyu Li, The Chinese University of Hong Kong, Shatin, Hong Kong, Wen Wen, Keongtae Kim

In the post-ChatGPT era, startups are increasingly entering the AI application market to launch their own bot services, seeking to attract investors and achieve long-term success. However, the majority of AI products and services being released appear to be of poor quality and fail to achieve product-market fit. This may even negatively impact investor and customer perceptions of these startups, given the increasingly crowded market. We study AI marketplace adoption to understand startups' strategies for launching AI complement products and their subsequent performance outcomes. Our preliminary findings suggest that launching AI bot services may actually harm funding performance.

TD10

Building A Level 3 A312

Digital Platform Economics and Strategy

Contributed Session

Contributed

Chair: Robin Ng, Mannheim, Mannheim, N/A, Germany

1 - Coexistence of Competition? Platform Entry Effect on Third-Party Sellers on Retail Platforms

Zohre Yahyaee, the University of Kansas, Lawrence, KS, United States, Anurag Garg, Arvind Tripathi

Hybrid e-commerce platforms—where marketplaces operate as both retailers and hosts to third-party (3P) sellers—have reshaped online competition. While platform entry can benefit or harm 3P sellers, empirical evidence remains inconclusive. This study examines the impact of first-party (1P) entry by JD.com on 3P seller performance, with a focus on product popularity. Using proprietary JD.com data and a Difference-in-Differences design with SKU- and day-level fixed effects, we find that 1P entry expands market size and increases 3P sales, but through different channels. For popular products, 1P presence reduces consumer uncertainty and raises conversion without increasing product views. For unpopular products, it boosts visibility but not sales. Prices remain stable across both segments. These findings clarify the heterogeneous effects of platform entry, highlighting the role of baseline demand and platform structure. The study offers strategic insights for platform operators, third-party sellers, and policymakers navigating competition in hybrid marketplaces.

2 - Supply, Demand OR Policy Driven? An Empirical Examination of Data Center Location Strategies in the United States

Chen Tian, Emory University, Atlanta, GA, United States, Ramnath Chellappa, Jonathan Gomez Martinez

The growth of cloud-based IT has fueled the rapid expansion of data center construction, raising questions about the determinants of digital infrastructure location. We examine the importance of market fundamentals—specifically supply-side and demand-side factors—as well as local government tax policies in influencing data center entry across U.S. counties. Using a panel dataset from 2000 to 2023 and a staggered difference-in-differences design, we find that state-level tax incentives significantly increase data center entry, particularly for public cloud and large colocation facilities. Additionally, data center entry continues to be shaped by supply and demand factors, including proximity to the information industry, land costs, and broadband access. However, time-varying results suggest that improvements in network latency and cloud technology have diminished the importance of demand-side considerations in data center location decisions. Our findings suggest a growing decoupling between data center location and digital demand, reflecting broader shifts driven by technological change and policy interventions.

3 - Investigating the Impacts of IT Outages on Firms' Service Operations In Digital Platforms

Huai-Tzu Cheng, Montana State University, Bozeman, MT, United States

IT outages create major challenges for firms, causing economic and operational setbacks and challenging information system resilience. These outages lead to the prevention of normal operations, lost revenue, reduced customer trust, and higher costs from downtime and recovery efforts, weakening a firm's competitive position in the market. Operationally, unexpected outages cause service delays and strain resources. Customer service is key to managing these impacts, but traditional support channels often get overwhelmed during the crisis, pushing firms to use digital platforms like X.com, Facebook, or Instagram for real-time customer communication. During crises, more firms increasingly rely on these platforms to connect and engage with customers and provide support. This shift requires fast responses and careful handling of misinformation to prevent negative word-of-mouth.

This study analyzes recent IT outages to quantify their impacts on customer service operations and evaluate information systems resilience, focusing on operational delays, customer service challenges, and recovery strategies. In this study, we find that customer complaints increase significantly during outages, and firms' slow responses worsen dissatisfaction. These findings advance academic understanding of IT outage impacts and provide practical strategies for firms to manage disruptions, enhance a firm's customer service, and optimize service recovery during crises in digital platforms, ensuring operational stability and customer satisfaction. This study contributes to the literature on information systems, crisis management, and service operations.

4 - Competition Through Recommendations

Robin Ng, Mannheim, Mannheim, Germany

Consumers rely on platform recommendations when purchasing products. I examine a platform's decision to design recommender systems to be informative of value. Informative recommendations generate two effects: screening effect, where lower-quality firms exit the platform; and ranking effect, among remaining firms, consumers interact more with higher-quality firms. Screening benefits higher-quality firms while ranking benefits only the highest-quality firms. Both effects improve consumer surplus, attracting more consumers to the platform. However, the ranking effect intensifies price competition as firms want to provide more value to attract demand. Hence, a platform designing its recommender system balances between stimulating price competition to attract consumers and per-transaction ad-valorem fees from firms. When platforms compete, recommendations become more informative when the competitor has a more informative recommender system. I connect these results to the evolution of recommender systems, showing how regulatory policies and the competitive environment lead platforms to prioritise trust-building or engage in enshittification.

TD11

Building A Level 3 A313

TSL Data-driven Research Challenge (Award Session 1)

Award Session

Transportation Science and Logistics (TSL)

Chair: Hai Wang, Singapore Management University, Singapore, Singapore

Co-Chair: Lei Zhao, Tsinghua University, Beijing, N/A

1 - Data-Driven Optimization for Meal Delivery: a Reinforcement Learning Approach for Order-Courier Assignment and Routing at Meituan

Ramon Auad, Amazon, Santiago, Chile, Tomas Lagos, Felipe Lagos

The rapid growth of online meal delivery has introduced complex logistical challenges, where platforms must dynamically assign orders to couriers while accounting for demand uncertainty, courier autonomy, and service efficiency. Traditional dispatching methods, often focused on short-term cost minimization, fail to capture the long-term implications of assignment decisions on system-wide performance. This paper presents a novel hybrid framework that integrates reinforcement learning with hyper-heuristic optimization to improve sequential order assignment and routing decisions in meal delivery operations. Our approach combines n-step SARSA with value function approximation and a multi-armed bandit-based hyper-heuristic incorporating seven specialized low-level heuristics. Our approach explicitly models the evolving system state, enabling dispatching policies that balance immediate efficiency with future operational performance. By employing scalable linear value function approximation, we enhance policy learning in high-dimensional environments while maintaining generalization across states and actions.

Using real operational data from the food delivery platform Meituan, we develop a comprehensive simulation environment that captures order dynamics, courier behavior, and service times. Through extensive computational experiments, we demonstrate that our framework significantly outperforms traditional benchmark policies, achieving 12% cost reduction through strategic order postponement. Our results reveal that the largest improvements occur during high-demand periods with courier shortages, and that a 10% increase in courier availability yields greater benefits than algorithmic improvements alone.

The proposed methodology effectively balances immediate operational efficiency with long-term performance, while providing valuable insights for meal delivery platforms regarding courier fleet management and order assignment strategies.

2 - Dispatch OR Hold? An Offline Learning and Online Optimization Approach for on-Demand Delivery

Long He, George Washington University, Washington, DC, United States, Yihua Wang, Zhengling Qi, Stefan Minner

In large-scale on-demand food delivery systems, dynamic order dispatching needs to balance immediate dispatch and order holding amid competing objectives, including delivery efficiency, service timeliness, and courier utilization. While holding orders may improve route consolidation, it also increases the risk of delayed delivery. Real-time decisions must determine both when to dispatch and how to group orders for efficient delivery.

We propose a hierarchical offline learning and online optimization framework that integrates inverse optimization with deep reinforcement learning. The framework first determines which orders to dispatch immediately and which to hold, then consolidates the dispatched orders by solving a multi-objective weighted set covering problem. To estimate the trade-off weights between objectives, we develop a cutting-plane-based inverse optimization method that learns from real-world dispatch data. The inferred cost structure is embedded in a Markov decision process, and a dispatch policy is trained with Proximal Policy Optimization to optimize long-term performance.

Empirical results on real-world data from a major food delivery platform show that the learned policy significantly improves operational outcomes, reducing travel distance and courier usage while improving on-time delivery. Analysis reveals key behavioral insights: strategic order holding increases grouping opportunities and improves delivery efficiency, particularly in periods or regions with low order arrival rates. However, long-distance orders are less suited for holding due to limited consolidation potential and higher delay risk.

3 - The Meal Delivery Routing Problem with Courier Behavior and Graph Neural Networks

Niels de Jong, University of Groningen, Groningen, Netherlands, Ilke Bakir, Kees Jan Roodbergen

We address the Meal Delivery Routing Problem (MDRP) faced by online food delivery platforms, where customer orders arrive dynamically and must be assigned to autonomous couriers. Each courier picks up orders from merchants and delivers them to customers. The platform must handle several challenges: couriers have varying behaviors and capabilities; courier-to-order assignment decisions must balance immediate and future costs; decisions must be made in seconds; multiple orders can be efficiently bundled; some assignments can be delayed; and there is inherent uncertainty in food preparation times, travel speeds, and order delivery locations. To address these challenges, we develop a scalable, polynomial-time heuristic using Graph Neural Networks (GNNs) to estimate future assignment costs in a rolling horizon assignment framework. We test our approach using a modular, open-source digital twin that simulates millions of deliveries in seconds and is validated on real data from Meituan. Our GNN-based heuristic is trained using Twin Delayed Deep Deterministic Policy Gradient (TD3). Results show our method reduces average delivery time (click-to-door) by 11.6% and order rejections by 2% compared to a myopic benchmark. It also adapts to varying courier behaviors and changing spatio-temporal demand, enabling real-time, efficient decision-making for food delivery and related applications like ride-sharing.

4 - Mixed-Type Courier Dispatch for Online Food Delivery Platforms

Junlin Chen, Tsinghua University, Beijing, China, People's Republic of, Manxi Wu, Chiwei Yan, Zhanhao Zhang

Many online food delivery (OFD) platforms employ a combination of full-time couriers, who operate on fixed schedules, and part-time couriers, who enter and exit the system dynamically. Despite their distinct operational characteristics, existing dispatch policies often overlook how to best allocate orders between these two groups. Using a large-scale dataset from Meituan—China's largest OFD platform, we identify systematic differences in courier behavior and performance, and develop a mathematical optimization model to improve dispatch efficiency. Our model reveals a spatial cutoff distance to guide order allocation: full-time couriers are prioritized for deliveries within this range, while part-time couriers handle orders beyond it. We develop analytical models for both single-cluster and multi-cluster settings, where each restaurant cluster has its own order arrival rate, distribution and dedicated full-time fleet, and part-time couriers operate across clusters. In both cases, we derive closed-form solutions for the optimal cutoff distance that maximizes system throughput. Furthermore, we allow the platform to allocate its total full-time couriers across clusters, formulate the resulting joint courier allocation and dispatching problem as a convex program, and provide an efficient algorithm. Our numerical experiments calibrated by real-world data confirm the practical value of our approach, noticeably improving on-time performance and throughput compared to benchmark policies.

TD12

Building A Level 3 A314

Coordinated Planning in Complex Transport Networks

Contributed Session

Contributed

Chair: Bhargav Ganguly, Amazon

1 - Designing Platooning Hub Networks Under Cost Uncertainty

Elif Zeynep SERPER, TED University, Ankara, Turkey, Sibel Alumur Alev

Platooning offers promising cost and environmental benefits for full truckload transportation by enabling trucks to travel in close convoys using vehicle-to-vehicle communication. This study addresses the design of hub networks that support platoon formation and dissolution, with the goal of transporting truckload shipments from their origins to destinations within the required delivery time while minimizing total costs.

We first develop a deterministic optimization model to identify optimal hub locations and routing decisions under known cost parameters. Each shipment can either be delivered directly or routed via two hubs to benefit from platooning. To account for real-world uncertainties in platoon cost savings, we then propose a two-stage

stochastic programming model. The Sample Average Approximation (SAA) method is employed to solve the stochastic model efficiently.

Using real-world data from 1,253 shipments among 39 U.S. cities, our computational results show that platoon-enabled networks significantly reduce transportation costs and improve delivery performance. The stochastic model provides improved performance under uncertainty, offering quantifiable gains over deterministic planning. These results offer practical insights for designing reliable and cost-effective platoon-based freight systems.

2 - A Game-Theoretic Model of Code-Sharing Between Marketing and Operating Airlines Based on Nash Bargaining and Second-Order Cone Programming

Bing-Hsin Sun, National Tsing Hua University, Hsinchu, Taiwan, Yu-Ching Lee, Chun-Han Wang

As the airline market matures, airlines have shifted from pure competition to strategic cooperation. Code-sharing is a common collaborative strategy that allows a marketing airline to sell seats on flights it does not operate, while the operating airline conducts the actual service. The revenue from such flights is distributed based on a profit distribution parameter. This strategy helps enhance network coverage, reduce operational costs, and improve overall market efficiency. Under a code-sharing arrangement, airlines compete on flight frequency while simultaneously collaborating on profit and market share distribution, forming a complex, multi-player game with strategic interdependence. Despite its strong practical relevance, current literature lacks a mathematical model that integrates both code-sharing agreements and frequency competition. To address this gap, this study proposes a game-theoretic model formulated as a second-order cone programming (SOCP). Using the Karush-Kuhn-Tucker (KKT) conditions, we derive a system of primal, dual, and complementarity slackness constraints. The model's objective is to maximize the product of airline profits, thereby obtaining a stable Nash bargaining solution (Nash, 1950). We further analyze how different profit distribution parameter affect airline profits, flight frequencies, passenger volumes, and the operational characteristics of code-sharing segments. Empirical results show that code sharing not only increases total industry profit and aircraft utilization but also allows airlines to serve more passengers without increasing the number of flights. The proposed model thus offers practical decision support for airlines formulating code-sharing strategies. Moreover, the findings confirm that code sharing can significantly enhance both profitability and service quality in the airline industry.

3 - Optimization of Hydrogen Refueling Facility Locations Under Dynamic Market Share Scenarios Leveraging Large Language Models

Mina Kim, Georgia Institute of Technology, Atlanta, GA, United States, Benoit Montreuil

The decarbonization of freight transportation has accelerated interest in hydrogen as a sustainable fuel alternative for heavy-duty trucks. A key barrier to large-scale adoption is the lack of adequate hydrogen refueling infrastructure. Static assumptions regarding vehicle market shares fails to capture the evolving nature of technology adoption and policy-driven market shifts. This study proposes a novel framework for optimizing hydrogen refueling facility placement by leveraging Large Language Models (LLMs) to generate dynamic market share scenarios for hydrogen fuel-cell electric heavy-duty trucks (FCEV HDTs). In contrast to conventional models, our approach anchors facility deployment on the Physical Internet (PI) regional hubs, which serve as strategic nodes in a hyperconnected logistics network. LLMs are used to synthesize plausible regional adoption patterns that reflect future uncertainty and heterogeneity in FCEV HDT uptake. These scenarios are integrated into an optimization model that selects hydrogen refueling facility locations to minimize emissions and cost while maximizing network coverage. Leveraging real-world freight data and the PI network, our method reveals deployment strategies that offer new perspectives beyond the reach of static models. The results highlight the potential of combining LLM-driven market forecasting with infrastructure planning to support resilient and future-ready hydrogen ecosystems.

4 - Consolidation of Urban Freight Delivery and Public Transit Systems

Mohammadhosein Pourgholamali, Purdue University, West Lafayette, IN, United States, Mohammad Miralinaghi, Samuel Labi

The demand in urban freight delivery systems is increasing, and freight agencies try to expand their delivery fleet to accommodate this demand. This is due to the expansion of e-commerce sales and online shopping over the past decades. The expansion in logistics fleet and last-mile delivery operation, however, caused many negative impacts on the community, such as traffic congestion. Therefore, having a more efficient and sustainable urban freight delivery system is necessary. Regarding the public transit, however, the demand has decreased. Therefore, the loss of ridership in recent years caused a loss of revenue for transit agencies and made the transit infrastructure underutilized. Public transit networks tend to be extensive, covering a large fraction of any metropolitan area, thereby offering wide coverage for small-freight collection and delivery. This study proposed a consolidated transportation system (CTS) in which the public transit fleet is used to transport passengers and packages in an urban area. A mathematical decision-making framework is developed to identify optimal transit lines, transit stations, and service zones that should be used to form a CTS. The analytical framework is formulated as an optimization problem that minimizes the total delivery cost and is solved by an Adaptive Large Neighborhood Search algorithm. The numerical results show promising benefits for freight agencies by saving delivery costs and for transit agencies by generating more revenue, respectively. Furthermore, the results show the important roles of different cost components in the feasibility of the CTS.

5 - Dynamic Fleet Allocation for Last-Mile Delivery Networks

Bhargav Ganguly, Amazon, Seattle, WA, United States, John Phelix, Andre Snoeck, Jin Ye, Daniel Merchan

Vehicle capacity represents a critical constraint in last-mile delivery optimization, as both physical cargo limitations and battery capacity influence route productivity. While increased vehicle capacity offers greater delivery potential, it increases the cost of operations, both in direct operational expenses and indirect logistical challenges such as urban parking constraints. We present a novel dynamic fleet allocation algorithm that optimizes the distribution of vehicle resources among Delivery Service Providers (DSPs) based on geographical service areas, demand forecasts, and anticipated weather. The proposed framework demonstrates improved vehicle utilization and cost efficiency compared to static allocation methods, particularly in the context of electric vehicles where effective capacity is highly temperature dependent.

TD13

Building A Level 3 A315

Resilient Logistics and Adaptive Strategies

Contributed Session

Contributed

Chair: Simanta Barman, University of Minnesota

1 - Stochastic Shockwave Dynamics in First-Order Traffic Flow

Yanlin Zhang, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Alireza Talebpour

First order continuum models, including the Lighthill-Whitham-Richards (LWR) model, have been studied extensively in the literature. The traditional formulation of these models assumes deterministic inputs, leaving a gap in our ability to describe how random upstream fluctuations propagate and interact with downstream traffic. In this work, we propose a novel framework for solving the LWR model with stochastic boundary conditions, where the upstream density evolves as a Markov process with only upward jumps, representing shock-inducing fluctuations.

Assuming a strictly concave fundamental diagram, this setting leads to a conservation law, where solution evolves according to a closed-form kinetic equation. This equation describes the statistical dynamics of traffic density over space and time, accounting for shockwave dynamics in a unified way.

We validate our approach through numerical experiments under both triangular and other smooth fundamental diagrams, and compare the statistical shock dynamics predicted by our kinetic model against high-resolution results from multiple discretization methods, including the Cell Transmission Model (CTM) and other Godunov schemes. Results show that our framework accurately reproduces shockwave statistics with significantly reduced computational cost and without the need for space-time discretization.

This framework provides an analytical tool for modeling traffic under uncertainty and opens the door to more robust strategies for congestion prediction and control, particularly in environments with varying driving behaviors.

2 - Multicommodity Trade, Tariffs, and Rerouting

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

The concept of “equilibrium” is fundamental to many disciplines, including economics and operations research. The theory of variational inequalities continues to provide a rich methodological framework for the formulation, analysis, and solution of a wide range of equilibrium problems with many recent applications, because of their relevance, including policies associated with the trade of commodities. In this paper, we add to the literature on multicommodity trade by constructing a spatial price equilibrium model with tariffs that allows for the investigation of rerouting of commodities to evade tariffs, as has been happening for several years now, along with “trade wars.”

3 - Causal Evidence on Economic Consequences of Maritime Shipping Disruptions

Shahnawaz Akhtar, IESE Business School, Barcelona, Spain, Mihalios Markakis, Geert Mesters

This article provides the first comprehensive evidence on the causal impact of maritime shipping delays on firm-level output and supply chain networks. Using a novel instrumental variable approach based on weather-induced port disruptions, we isolate the causal effect of delays on firm sales, leveraging a unique dataset that integrates firm financial data with detailed shipping records. We find that firms experiencing severe shipping delays suffer significant output declines. Moreover, these shocks propagate beyond the initially affected firms, leading to measurable contractions in output among their upstream suppliers and downstream customers. Industries heavily reliant on maritime trade, such as manufacturing and wholesale trade, are particularly vulnerable. These findings offer critical insights into the economic risks posed by shipping disruptions and underscore the urgent need for resilient supply chain strategies to mitigate their impact.

4 - Optimizing Truck Parking Infrastructure Using a Bi-Objective Flow-Based Model

Aida Riahifar, University of Nebraska-Lincoln, Lincoln, NE, United States, Nathan Huynh, Li Zhao, Zhenghong Tang

The shortage of safe and adequate truck parking capacity poses a critical challenge to improving freight transportation safety across the United States—a challenge exacerbated by the enforcement of Electronic Logging Device (ELD) regulations and Hours-of-Service (HOS) requirements. These mandates compel drivers to stop within defined time intervals, yet limited access to designated rest areas increases crash risk, driver fatigue, and disrupts supply chain operations. This raises an essential research question: how can truck parking locations and capacities be optimally planned to address safety requirements while remaining financially feasible?

To answer this question, this study develops a bi-objective facility location model that jointly optimizes the placement and scale of truck parking infrastructure. The novelty of the approach lies in modeling crash risk as a spatial-temporal probabilistic flow and infrastructure cost as a reverse flow within a graph-theoretic flow network framework. A multi-objective optimization strategy, supported by scenario-based and sensitivity analyses, is used to evaluate trade-offs between safety and investment cost. The model is applied as a case study to Interstate 80 (I-80) in Nebraska—a heavily trafficked and strategically vital corridor for east-west freight movement in the U.S.

The findings identify high-priority investment zones and offer a decision-support framework for policymakers and logistics planners. By leveraging a data-driven and scalable approach, this framework enables strategic alignment of parking infrastructure with dynamic freight demand, enhancing safety, regulatory compliance, and the overall resilience of freight transportation networks.

5 - Analysis of Throughput Properties of A Max-Pressure Based Ramp Metering Control

Simanta Barman, University of Minnesota, Minneapolis, MN, United States, Abdullah Farabi, Michael Levin, Ali Hajbabaie

Ramp metering is proven to be an effective strategy for improving freeway traffic operations. Previous studies have focused on optimizing various objectives such as freeway throughput, speed, waiting time, and capacity utilization. While these methods effectively manage freeway operations in different scenarios, most ramp metering controls lack theoretical guarantees of performance. In this study, we propose a novel ramp metering control framework based on max-pressure principles, aiming to maximize freeway flow rate. We provide a mathematical proof that the proposed control can serve nearly all of the feasible demand that any control method can handle. Additionally, we present a simple, efficient algorithm for implementing the proposed control. Numerical and simulation results demonstrate the significant effectiveness of our approach. Numerical analysis shows that under realistic conditions, the proposed control can serve more than 99% of the maximum theoretical demand. In a case study, we implemented our Max-Pressure Ramp-Metering (MPR) control in a microsimulation environment and benchmarked its performance against no metering control and the state-of-the-art ALINEA method. Simulation results show that the proposed control outperforms both benchmarks, reducing delay by up to 56% and average travel time by up to 20%.

TD14

Building A Level 3 A316

Innovations in Mobility Services

Contributed Session

Contributed

Chair: Breno Serrano de Araujo, Technical University of Munich, Munich, Germany

1 - Optimizing Curb Space Management Through Dynamic Pricing

Tien Nguyen, University of Michigan-Dearborn, Dearborn, MI, United States, Armagan Bayram

With the growth of new mobility services and increased needs for goods delivery, curb spaces, and their use have evolved very rapidly. Especially after COVID-19, active travel has increased and delivery rates have skyrocketed, creating a great need to ensure safe and efficient curb management. With delivery trucks, ride-sharing vehicles, private cars, and pedestrians all needing curb space, cities need a better way to manage these areas to keep them safe and efficient. One way to handle this is by using flexible pricing for various curb uses that adjust based on demand and available curb spaces. Hence, in this study, we developed a pricing model that sets curb prices dynamically for different uses by considering the uncertain demand, to maximize the revenue cities earn while supporting a fair and organized use of curb space for everyone. Our model aims to create straightforward pricing strategies that city planners can put into action quickly. This approach helps cities adapt to the fast-changing needs of curb spaces, supporting smoother, safer, and more organized city streets for everyone.

2 - Estimating Consumer Willingness-to-Pay for Time Saving Using Maas: Contingent Valuation Method Approach

Stephen Park, University of Goettingen, Goettingen, Germany, Hyungbin Moon, Yasemin Boztug

Mobility as a Service (MaaS) platforms offer innovative solutions by integrating various transportation modes into a single, user-friendly service. While MaaS is widely recognized for its convenience and sustainability benefits, the social benefit of time savings—an essential metric for urban transportation efficiency and productivity—remains underexplored. This study addresses this gap by estimating consumers’ willingness to pay (WTP) for time savings on MaaS platforms. Using the Contingent Valuation Method (CVM), we analyze personal preferences towards different time savings from a diverse urban sample, incorporating covariates such as gender, age, income, residency, commuting status, and access to mobility services. Based on the analysis, various consumer clusters that have homogeneous WTP for travel route search and travel time savings are identified, allowing for segmentation of urban mobility market. Characteristics of different clusters are examined to determine if seamless integration of different mobility services or reducing searching time is more valued by different consumer segments. The results reveal that consumers place significant economic value on search time reduction, with frequent commuters and urban residents with limited mobility alternatives exhibiting a higher WTP. These findings provide valuable insights for policymakers and service providers, underscoring the economic benefits of streamlining route-planning interfaces and supporting the

development of pricing models and platform features that prioritize time-saving aspect of the platform. Looking ahead, future research will extend this framework to analyze WTP for actual travel time reduction, enabling a more comprehensive understanding of the value proposition of MaaS across all phases of urban mobility.

3 - Optimizing Customized Bus Lines Considering Users' Transfer Willingness Under Cooperative and Competitive Relationship Between Metro and Online Car-Hailing

Beibei Wang, School of Economics and Management, Southeast University, Nanjing, China, People's Republic of, Xinyi Qi

In the context of 'carbon peak' and 'carbon neutrality', coordinating individual travel demand through multi-modal transportation and guiding travelers towards new shared public transport options is increasingly important. In this paper, we analyze the competitive and cooperative relationship between online car-hailing (OCH) services and metro systems in Nanning, China, and conduct a survey among different types of OCH users. A mixed choice model that considers psychological latent variables is constructed to investigate OCH users' attitudes and cognitions toward customized buses (CBs). An improved adaptive Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is proposed to identify potential carpooling station sets, and a hybrid genetic-ant colony algorithm (GACA) is designed to solve bi-level programming model for CB line optimization. Additionally, a case study in Nanning indicates an 83.8% overall transfer rate from OCH users to CBs, with the optimized scheme achieving a 69.68% reduction in carbon emissions.

4 - Joint Optimization of Emerging Mobility Infrastructure and Services with Considerations of Accessibility and Fairness

Lindsay Graff, Carnegie Mellon University, Pittsburgh, PA, United States, Katherine Flanigan, Peter Zhang, Sean Qian

Cities are integrating new mobility modes into their transportation networks to improve route options and increase accessibility. Strategically co-locating mobility infrastructure and services is consequently of pivotal importance. This work introduces a bi-level, bi-objective mixed-integer program to jointly select the locations of bikeshare stations and microtransit service zones. The upper-level objective is a weighted sum of a system-level accessibility metric and the minimum value of the accessibility metric computed for any origin. The lower level solves the combined mode split and traffic assignment problem for a multimodal system with shared mobility. Since the relationship between the lower- and upper-level decision variables cannot be expressed in closed form, an iterated local search heuristic is used. A case study in Allegheny County, PA illustrates the benefit of joint optimization. When bikeshare and microtransit investments are optimized together, system-level accessibility increases by 4.0% relative to the transit-only baseline, and minimum-level accessibility rises by 8.7%. A key insight is that bikeshare and microtransit individually provide limited improvements for the most disadvantaged area, but their impact magnifies when deployed together. Experiments reveal diminishing marginal returns of the bikeshare budget on the combined objective value due to diminishing returns on the minimum-level accessibility metric. This suggests that the constraints of the existing public transit network limit the extent to which shared mobility can improve accessibility in the worst-off areas. Finally, Pareto optimization illustrates the importance of balancing system and fairness objectives to ensure that disadvantaged areas are not left behind in pursuit of broader system-wide gains.

5 - Optimizing a Ride-Hailing System with a Mix of on-Demand and Pre-Booked Customers Under Distributional Shift

Breno Serrano de Araujo, Technical University of Munich, Munich, Germany, Alexandre Jacquillat, Stefan Minner, Maximilian Schiffer

We consider a mixed-service ride-hailing system that offers customers the option to request a ride on-demand or to pre-book it in advance. For time-sensitive customers, pre-booking provides a service guarantee at a price premium. From the operator's perspective, pre-booking allows for planning ahead with higher certainty but may incur the duty to operate unfavorable trips that may even induce a shift in the demand distribution, e.g., in low-demand suburban neighborhoods. Against this background, we develop an optimization framework for a mixed-service ride-hailing system, allowing us to study the trade-offs between higher planning certainty and the rise of unfavorable rides due to shifts in the demand distribution. We propose a two-stage stochastic optimization formulation in which the first-stage problem consists of deciding which pre-booking requests to accept, while the second-stage problem involves assigning vehicles to requests and planning routes with uncertain on-demand requests. We present a sample average approximation formulation and a scalable solution algorithm that approximates the second-stage subproblems using a polynomial-time algorithm. We conduct experiments based on yellow taxi trip data from New York City. We show that greedily accepting all pre-booking requests leads to a 14.5% profit reduction compared to a purely on-demand baseline in environments with strong distributional shifts. In contrast, our proposed solutions lead to profit increases between 6.5% and 7.7%, in settings with weak and strong distributional shifts, respectively, while satisfying customer pick-up time windows. Additionally, we provide managerial insights to assist operators in making informed design choices to counteract the effects of distributional shifts.

TD15

Building A Level 4 A401

Urban Logistics Network Design and Operations

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Dipayan Banerjee, Loyola University Chicago, 220 W Illinois St, Chicago, IL, 60654, United States

1 - Service Network Design in Intra-City Express Delivery

Tirui Cao, Tsinghua University, Beijing, China, People's Republic of, Lei Zhao, Yixiao Huang, Alexandre Jacquillat, Alexandria Schmid

Intra-city express delivery is an emerging service that enables customers to send and receive packages in the same city within a short delivery window. Its time- and cost-efficiency relies on a dedicated intra-city service network, where packages are transported between and consolidated at a number of satellite stations located in the city. A key challenge in designing and operating such service networks lies in the limited docking capacity to unload/load packages at satellite stations, which often leads to congestion and delays. Besides increasing docking capacity at certain satellite stations, one may also reroute a subset of intra-city packages using the spare capacity of the inter-city service network (independently designed for inter-city express delivery). This mitigates congestion at the satellite stations and shortens the delivery time for the remaining packages, yet introduces extra detour and sorting time for the rerouted packages. We show in a theoretical model that, when these trade-offs are properly balanced, leveraging the inter-city service network can shorten average delivery time and reduce total operational cost. Further, we formulate the intra-city express delivery service network design problem as an integer programming model and explore two solution algorithms: a column generation-based algorithm and a multi-arc generation-based algorithm.

Based on real world instances from a major express delivery carrier in China, our preliminary numerical experiments demonstrate the scalability of the proposed algorithms and validate the value of leveraging the inter-city service network to improve service quality and reduce operational cost in intra-city express delivery.

2 - Stochastic Fleet Coordination in Disaster Relief: a Model for Integrated Evacuation and Supply Delivery

Mina Samiei Nasab, University at Buffalo, Buffalo, NY, United States, Rajan Batta

This research addresses a dynamic transportation planning problem in the context of post-disaster humanitarian logistics, an area that lies at the heart of the Transportation Science and Logistics (TSL) domain. Specifically, we develop an integrated model to coordinate a heterogeneous fleet of vehicles to serve victims with evolving and time-sensitive needs. Victims are classified into two groups: those requiring immediate evacuation (Category A) and those initially needing critical supply delivery (Category B). However, Category B victims can deteriorate with time, creating cascading demand on already limited evacuation resources. Most existing models treat evacuation and delivery as separate problems, overlooking the synergistic potential of joint coordination. Our research bridges this gap by introducing a unified optimization framework where different vehicle types work in coordination to handle both evacuation and delivery operations, expanding operational reach and enabling more flexible, adaptive operations. The core problem involves the simultaneous routing of multiple vehicle types, selection of service points, and real-time reassignment based on endogenous uncertainty in service times and victim deterioration.

3 - Joint Bus Trip Scheduling and Vehicle Assignment: A Scalable Iterative Framework for Time-Dependent Travel Times

Yuhan Tang, Massachusetts Institute of Technology, Cambridge, MA, United States, Charlie Liu, Chonghe Jiang, John Attanucci, Haris Koutsopoulos, Jinhua Zhao

This research addresses the inefficiencies in traditional bus scheduling systems that rely on fixed runtime estimates, which fail to capture delay propagation—where delays on one trip adversely affect subsequent trips served by the same bus—and passenger demand variability. To overcome these shortcomings, we develop an integrated optimization model that simultaneously determines bus trip departure times and bus assignments using a mixed-integer quadratic programming framework. The model also incorporates robust optimization to handle travel time uncertainty, ensuring effective scheduling under real-world operational conditions.

Our approach features a network flow formulation for bus assignment that yields a notably tight linear relaxation, enhancing both solution quality and computational efficiency. To improve scalability and tractability, we employ Benders decomposition to decouple the problem into two stages: a trip scheduling phase focused on minimizing total passenger waiting time, and a bus assignment phase formulated as a minimum path cover problem in an acyclic graph.

Numerical experiments and a case study on Chicago Transit Authority's Route 22 demonstrate that the integrated model significantly reduces passenger waiting times and improves operational efficiency compared to conventional methods. Academics benefit from an advanced theoretical framework that integrates delay propagation and robust optimization. Transit practitioners gain a data-driven tool for enhancing service reliability, while policy makers receive quantitative insights supporting sustainable urban mobility and informed resource allocation.

4 - Service Network Design of Modular Vehicles on General Road Networks Considering Passengers' Turning Choices

Xiaolei Wang, Tongji University, Shanghai, China, People's Republic of, Jiahui Yu, Lei Guo, Zhibin Chen, Zhe Liang

The emergence of modular vehicles (MVs) opens the avenue to an entirely new urban transit system, in which a fleet of modular vehicles circulating within the road network constitute a dynamic MV service network, and transit passengers can freely and stop-lessly choose where they go at each intersection by transferring to the appropriate pod. Provided feasible MV circulation patterns on the road network, this paper develops a novel mathematical model to depict passengers' turning choices at user equilibrium in the MV service network. We establish the equivalence of this model to a linear optimization problem, which can be further transformed into a multi-commodity network flow problem on a revised network. Taking passengers' turning choices as the lower-level problem, we then propose a bi-level problem to optimize the MP circulation pattern. We show that this bi-level problem can be transformed into a single-level problem with bi-convex property, and employ a multi-start alternating minimization (AM) algorithm to solve it by iteratively solving two convex subproblems. The convergence of the AM algorithm is theoretically established, and tailored algorithms are developed for solving each subproblem efficiently in large-scale networks. Numerical experiments are conducted to validate the model and algorithm, and demonstrate the computational efficiency on large-scale networks.

5 - Optimal Batching and in-Building Delivery Routing with Capacitated Residential Parcel Lockers

Dipayan Banerjee, Loyola University Chicago, Chicago, IL, United States, Ignacio Erazo

In this work, we optimize the operations of a parcel delivery firm that has access to a residential parcel locker (RPL) in the lobby of a high-rise building. We assume that the number of parcels to be delivered on a particular day exceeds the number of available RPL compartments; parcels that cannot be delivered to the RPL must be delivered to residents' doorsteps on upper floors of the building. Given certain building features, such as floor layouts and elevator waiting time distributions, we seek to determine the optimal subset of parcels to deliver to the RPL. For the remaining parcels, we seek to optimize the driver's travel route within the building; when the number of non-RPL packages exceeds the driver's carrying capacity, we additionally seek to determine an optimal separation of the non-RPL packages into batches. Our objective is to minimize the expected time to deliver all of the building's parcels. Under mild assumptions, we first derive a polynomial-time algorithm for determining the optimal sequence of floor visits given a set of parcels to deliver at residents' doorsteps. We then leverage this algorithm to develop a branch-and-price solution approach capable of solving realistically sized instances of the full problem to optimality.

TD16

Building A Level 4 A402

Striking the Balance of Innovation and Responsible Ethical AI

Panel Session

Practice Track

Co-Chair: Archi Pandurangi, ISpot TV, Alpharetta, United States

1 - Moderator Panelist

Archi Pandurangi, ISpot TV, Alpharetta, GA, United States

As artificial intelligence reshapes industries and decision-making at scale, the tension between rapid innovation and responsible governance grows ever sharper. This panel explores how leaders and technologists can drive cutting-edge advancements while upholding ethical standards that protect individuals, communities, and democratic values. From algorithmic transparency and bias mitigation to compliance frameworks and long-term societal impacts, panelists will examine real-world trade-offs and decision points. With perspectives spanning enterprise analytics, product development, and regulatory oversight, the discussion will highlight pragmatic approaches for embedding ethics into scalable AI systems—ensuring that innovation doesn't outpace accountability.

2 - Panelist

Ritu Sharma, Equifax, Alpharetta, GA, United States

TD17

Building A Level 4 A403

AI-Powered Data-Driven Decisions Throughout the Consumer Lending Lifecycle

Panel Session

Practice Track

Co-Chair: Jim Bander, Experian, Tempe, 85284-2558, United States

1 - Moderator Panelist

Jim Bander, Experian, Tempe, AZ, United States

This panel discussion will explore how artificial intelligence and advanced data Description: analytics are revolutionizing decision-making across the entire consumer lending lifecycle. Experts from Experian, JPMorgan Chase, and Porsche Financial Services will share their experiences in leveraging AI and diverse data sources, including traditional bureau data and innovative alternative information, to optimize strategies from underwriting and loan pricing to collections risk management. The panelists will delve into the practical applications of machine learning, algorithms, and other AI-driven techniques to enhance efficiency, mitigate risks, and improve outcomes in today's complex and evolving lending landscape.

2 - Panelist

Sarah Zeis, Porsche Financial Services, Atlanta, GA, United States

TD18

Building A Level 4 A404

Military and Security 1

Contributed Session

Contributed

Chair: Murat Balci, Old Dominion University, 5115 Hampton Blvd, Norfolk, VA, 23529, United States

1 - Optimal Pure Strategies for a Discrete Search Game**Thuy Bui, Rutgers University, New Jersey, NJ, United States**

Consider a two-person zero-sum search game between a Hider and a Searcher. The Hider chooses to hide in one of n discrete locations (or “boxes”) and the Searcher chooses a search sequence specifying which order to look in these boxes until finding the Hider. A search at box i takes t_i time units and finds the Hider—if hidden there—independently with probability q_i , $i = 1, \dots, n$. The Searcher wants to minimize the expected total time needed to find the Hider, while the Hider wants to maximize it. It is shown in the literature that the Searcher has an optimal search strategy that mixes up to n distinct search sequences with appropriate probabilities. This paper investigates the existence of optimal pure strategies for the Searcher—a single deterministic search sequence that achieves the optimal expected total search time regardless of where the Hider hides. We identify several cases in which the Searcher has an optimal pure strategy, and several cases in which such optimal pure strategy does not exist. An optimal pure search strategy has significant practical value because the Searcher does not need to randomize their actions and will avoid second guessing themselves if the chosen search sequence from an optimal mixed strategy does not turn out well.

2 - A Prioritized 2-D Packing Model for Military Combat Loading**William Kirschenman, North Carolina State University, Raleigh, NC, United States, H. Sebastian Heese, Michael Kay, Russell King, Brandon McConnell**

Rapid and accurate combat loading of personnel and equipment is critical to success in large-scale, contested military operations. Existing tools like the Integrated Computerized Deployment System (ICODES) Single Load Planner assist with generating load plans but rely heavily on manual item placement and constraint checks performed by the user interface’s built-in “agents.” This research introduces a prioritized two-dimensional bin packing (2D-BPP) approach to automate and optimize these decisions. We formulate the problem as a mixed-integer linear program (MILP) minimizing a weighted rectilinear distance objective capturing multi-level tactical priorities for item placement relative to each other and access points. The MILP model integrates essential constraints including non-overlap, fixed obstacle avoidance, minimum item separation, and critically, global load balancing via center-of-gravity limits.

While smaller instances are solvable directly using commercial solvers (e.g., Gurobi), we utilize a two-stage framework. Stage 1 uses a sliding-window decomposition method for an initial plan optimized for the prioritized distance objective. Stage 2 applies a post-processing MILP adjustment to meet load balancing constraints. This MILP performs minimal modifications targeting center-of-gravity limits for load balance, preserving Stage 1 tactical priorities. Compared to alternatives like in-stride penalization during Stage 1’s iterative subproblem solves, this post-processing MILP adjustment within the two-stage framework is the most promising approach for effectively balancing competing objectives.

Results with realistic military equipment data show this two-stage decomposition framework rapidly generates high-quality, balanced, tactically sound load plans, significantly improving upon current methods. We analyze performance trade-offs between the exact MILP and this two-stage approach.

3 - Comparing Milp Formulations for Medevac Mission Planning**Armin Fugenschuh, Brandenburg University of Technology, Cottbus, Germany**

Rapid and reliable air- or ground-based medical evacuation (MEDEVAC) remains a core operational challenge in military missions. We introduce and compare two mixed-integer linear programming (MILP) models that automate MEDEVAC dispatching while explicitly accounting for escort vehicles, mission safety, and limited resources.

The first model is a **time-expanded network** formulation that discretizes the planning horizon into successive layers, tracking vehicle movements, patient transfers, and escort assignments on a replicated battlefield graph. The second, a **path-and-time** formulation, couples binary routing decisions with continuous departure and arrival times, thereby avoiding full time discretization while preserving all precedence and risk constraints.

A common set of benchmark instances will evaluate runtime, solution quality, and scalability, revealing the conditions under which each formulation excels or struggles. Beyond this comparative analysis, we outline avenues for integrating multicriteria objectives—such as balancing safety and speed—and for incorporating real-time information. These insights aim to guide practitioners in selecting and tuning optimization engines for next-generation, AI-assisted MEDEVAC planning tools.

4 - Strengthening Secure Communications: Addressing Operational Risks in Messaging Platforms**David Gurzick, Hood College, Frederick, MD, United States, Jonathan Shaw Wood, Gerard Werner**

Secure messaging applications like Signal offer end-to-end encryption, yet recent events have exposed significant operational risks when these tools are used improperly. A high-profile case involving a senior US official’s misuse of Signal highlighted how even strong encryption cannot prevent operational security failures without strict access controls. Drawing from these vulnerabilities and lessons learned, we propose a novel framework for secure communications built around three key principles: (1) isolated session-based communications, (2) ephemeral keying, and (3) continuous real-time access validation. In this model, each communication session forms a temporary, encrypted peer-to-peer tunnel established only after multi-factor authentication and dynamic authorization checks. Encryption keys are generated for single use and destroyed immediately after the session ends, leaving no persistent access points. Continuous endpoint monitoring ensures that any change in authorization status results in instant revocation of access, even during active sessions. Unlike traditional messaging platforms, this approach ensures that unauthorized individuals cannot join, monitor, or retroactively access sensitive exchanges. The design addresses gaps exposed by incidents like the Signal case, providing stronger operational resilience by embedding zero trust and dynamic control into the core of session management.

5 - Are More Competitive Supply Chains More Resilient? Evidence from Defense Industrial Base (DIB) Supply Chains**Yiseon Choi, The Ohio State University, Columbus, OH, United States, Noah Dormady, Ruhaimatu Abudu**

This paper explores the resilience of supply chains within the Defense Industrial Base (DIB), a critical infrastructure sector designated under Presidential Policy Directive 21 (PPD-21). We surveyed domestic and international firms involved in the supply chains of critical materials such as small arms munitions and eight DIB-critical metals, several of which are classified as strategic or rare. We collected data on resilience tactics (e.g., input substitution, resource isolation) employed by firms in response to disruptions and assessed the economic avoided losses that would have occurred in the absence of these tactics using a unique survey methodology. Through firm and tactic-level analyses, we evaluate the cost-effectiveness of resilience tactics and compare the severity of supply chain disruptions. This paper takes a novel approach to explore how firm size influences resilience and traces the pathways through which input substitutability affects firm resilience. We use Structural Equation Models (SEM) to estimate both the direct and indirect effects of having a larger pool of input suppliers on the magnitude of losses a firm can avoid by leveraging resilience tactics. Specifically, we trace the pathway from the size of the supplier pool through the expected duration of the disruption, and ultimately to the firm’s resilience outcome. The findings show that access to a larger pool of potential suppliers is associated with longer disruption durations, which in turn allows firms to avoid greater losses, particularly when operating under fixed-price supply contracts.

6 - Army Officer Rebranching: A Multicriteria Optimization Model for Rebalancing Officer Inventories**Matthew Ferguson, United States Army Human Resources Command, Elizabethtown, KY, United States**

The U.S. Army’s officer career fields are divided into branches with varying personnel needs; some require a surplus of junior officers, while others need more senior officers. Historically, the Army corrected these imbalances through a branch detailing process. However, the recent transition to a voluntary branch detail system has diminished the effectiveness of this approach, leading to projected mismatches in officer inventory that could negatively impact professional military education, career development and unit readiness. To address these challenges, the Army developed the Officer Rebranching Process (ORP) allowing an opportunity for junior Army officers another opportunity to voluntarily transfer branches as junior officers. The Army elicits ordinal and possibly incomplete preferences from junior officers eligible to transfer branches over a set of branches that are projected to require additional officers, and capacities for each branch, as well as limitations on transfers, are defined. We present the development and implementation of an optimization model supporting the ORP. The model employs a multicriteria lexicographic framework from the

literature to balance Army readiness requirements with officer preferences. This approach balances several objectives defined by stakeholders and resulted in rapid stakeholder agreement. The model was initially developed and utilized in 2024 and a revised version of the model used in 2025 to voluntarily reassign Army officers among branches while accounting for both Army readiness requirements and individual officer preferences.

7 - Time-Constrained Sample Average Approximation for Port Selection in Lateral Transshipment with Stochastic Travel Time

Lennon Conson, U.S. Air Force, Urbana, OH, United States, Brian Lunday, Matthew Robbins

This presentation introduces both a model and a time-constrained solution algorithm for port selection in lateral transshipment with stochastic travel times. In our problem, roll-on/roll-off equipment is dispersed across the continental United States. An emergency deployment order directs that a portion of the equipment be shipped to a single overseas seaport of debarkation (SPOD). The equipment must be moved first by rail to selected seaports of embarkation (SPOEs) and then shipped to the SPOD. Because only a limited number of SPOEs can be activated, a first-stage decision—which ports to open—must be made within a set amount of time.

We formulate the problem as a large-scale two-stage stochastic mixed-integer program that captures rail-to-ship synchronization, processing limits at each SPOE, and stochastic rail and ship travel times. Because the first-stage decision must be made within a limited amount of time before transshipment begins, we propose a Time-Constrained Sample Average Approximation (TC-SAA) framework. TC-SAA has two distinct phases: pre-deployment metamodeling and emergency-deployment. During the pre-deployment metamodeling phase, planners solve a representative set of SAA instances created with a space-filling design over plausible demand profiles. In the emergency deployment phase, the algorithm utilizes the first-phase metamodeling to choose an appropriate SAA sample size that allows for enough replications to reduce the predicted estimated optimal gap.

8 - Simulation-Supported Decision Analysis of Command Structures Under Contextual Uncertainty

Murat Balci, Old Dominion University, Norfolk, VA, United States

This research introduces a decision analysis framework supported by simulation to compare the effectiveness of centralized and decentralized command structures, under varying operational conditions. The framework is grounded in a rational inductive approach, combining conceptual modeling with discrete-event simulation to analyze decision performance under uncertainty.

Using Lanchester-based attrition modeling, the simulation explores scenario variations such as force imbalance, reinforcement timing, and decision latency. The outcomes are measured using two core decision objectives: mission effectiveness and resource efficiency. The framework is applied to compare how structural command constraints, unity of command versus decentralized responsiveness, affect outcome robustness in both deterministic and stochastic environments.

Rather than focusing on doctrinal correctness, this study treats command styles as decision architectures, each offering different trade-offs in flexibility, coordination, and risk exposure. The results provide structured insights into when and why particular decision frameworks outperform others depending on the operational context. The study contributes to the decision analysis literature by offering a generalized approach to evaluating organizational decision structures in complex, dynamic environments.

TD19

Building A Level 4 A405

Strategic Interactions in Service Systems

Invited Session

Applied Probability Society

Chair: Ran Snitkovsky, Collier School of Management, Tel Aviv University, Tel Aviv - Jaffa, Israel

1 - On the Value of Service Age Information

Ricky Roet-Green, Simon Business School, University of Rochester, Rochester, NY, United States

Sharing service age information with customers has become a common practice among electric vehicle (EV) charging companies. This information plays a crucial role in customers' waiting decisions, as charging times can range from 30 minutes to several hours. To reduce uncertainty and improve drivers' willingness to wait, many companies now disclose service age details. The vast majority of papers in the literature ignore service age information because they assume an exponential service time distribution. However, this Markovian service assumption, which is intended to simplify the analysis, is less common in real-life settings. In cases of general service distributions where the memoryless property does not hold, service age becomes crucial for helping customers better estimate their waiting time and plays an important role in their decision to join the queue. While recent research has highlighted that disclosing queue-length information to customers may improve system performance, the value of revealing the service age information remains unaddressed. Our work aims to close this gap.

2 - Preemption Pays: Optimal Preemptive Priority Pricing in M/G/1 Queues

Binyamin Oz, Hebrew University of Jerusalem, Jerusalem, Israel, Moshe Haviv, Elisheva Zur

In an M/G/1 queueing model where the mean residual service time is larger than the mean service time itself, preemption improves the performance of the system. One way to utilize this idea is by granting customers preemptive priority via a lottery. We derive the optimal probabilities of this lottery for any preassigned number of priority classes. In particular, the optimal probabilities define a truncated geometric distribution with a parameter which is a function only of the system's traffic intensity and the number of priority classes. Then we show how these probabilities can be implemented via the introduction of various prices across the priority classes, while the customers, who select and pay for a priority class, are engaged in a non-cooperative game among themselves. These probabilities define a mixed Nash equilibrium. Finally, we identify a range for the system's traffic intensity where, regardless of the number of priority classes, this equilibrium is unique and the (generalized) avoid the crowd (ATC) property prevails.

3 - Information Asymmetry In Queues with Strategic Customers

Shunan Zheng, University of Texas at Austin, Austin, TX, United States

In this work we study information asymmetry in an unobservable single-server queueing system. While system managers have knowledge of the true arrival rate, customers may lack this information and instead form arbitrary beliefs. We propose a three-tier hierarchy of information asymmetry with increasing levels of information disclosure. Within this framework, the effects of the belief distribution, which is assumed to be general with minimal restrictions, are analyzed in terms of equilibrium joining probabilities, revenue, and social welfare. Furthermore, strategies for information disclosure are proposed for system managers to regulate the queue.

4 - From Baseline to Boost: Optimizing and Re-Optimizing Price and Capacity in Response to Marketing-Driven Demand Growth

Levent Kocaga, Sy Syms School of Business, New York, NY, United States

Many service firms optimize pricing and capacity to serve customers who are sensitive to both price and congestion. In such settings, the endogenous demand is influenced by both price and capacity which leads to a complex equilibrium with no closed-form expression for the arrival rate. This challenge is further complicated when firms undertake marketing promotions, such as advertising, which expand the market size over time. In response, firms must re-optimize pricing, capacity, or both to effectively leverage this increased demand. We study this joint pricing and capacity problem in the context of a large-scale service provider modeled as an unobservable M/M/C (Erlang-C) queue. Customers have heterogeneous valuations and choose to purchase if their valuation exceeds the sum of the price and expected delay cost. This strategic behavior induces an endogenous equilibrium arrival rate, and the provider aims to maximize profit—defined as revenue minus linear capacity costs—based on this equilibrium. We distinguish between initial optimization, which targets a baseline market size, and re-optimization, which occurs post-promotion. Depending on operational flexibility, the provider may adjust both price and capacity or only one. Because exact equilibrium analysis is intractable, we use an asymptotic approach: letting market size grow while holding individual preferences fixed. This yields tractable, near-exact expressions for optimal decisions and profit. Our results show that

substantial gains from market expansion can be captured by adjusting capacity alone. We also account for marketing costs and provide guidance for designing near-optimal promotions that align with operational capabilities and profit goals.

TD20

Building A Level 4 A406

Learning in Dynamic Environments

Invited Session

Applied Probability Society

Chair: Moise Blanchard, Georgia Tech, Somerville, MA, United States

1 - Sequential Learning Under Distribution Shifts

Samory Kpotufe, Columbia University, New York, NY, United States

We consider sequential learning, e.g., ‘online learning’ and ‘bandits’, under distribution-shifts settings that interpolate between fully adversarial and stochastic (i.e. where observations are i.i.d.). The main questions of interest revolve around the types performance possible under relevant notions of performance, e.g., dynamic or adversarial regrets, and elucidating conditions on intervening distributions that affect such performance.

2 - Conformal Prediction Under Levy-Prokhorov Distribution Shifts: Robustness to Local and Global Perturbations

Julie Zhu, MIT, Cambridge, MA, United States, Liviu Aolaritei, Oliver Wang, Michael Jordan, Youssef Marzouk

Conformal prediction provides a powerful framework for constructing prediction intervals with finite-sample guarantees, yet its robustness under distribution shifts remains a significant challenge. This paper addresses this limitation by modeling distribution shifts using Levy-Prokhorov (LP) ambiguity sets, which capture both local and global perturbations. We provide a self-contained overview of LP ambiguity sets and their connections to popular metrics such as Wasserstein and Total Variation. We show that the link between conformal prediction and LP ambiguity sets is a natural one: by propagating the LP ambiguity set through the scoring function, we reduce complex high-dimensional distribution shifts to manageable one-dimensional distribution shifts, enabling exact quantification of worst-case quantiles and coverage. Building on this analysis, we construct robust conformal prediction intervals that remain valid under distribution shifts, explicitly linking LP parameters to interval width and confidence levels. Experimental results on real-world datasets demonstrate the effectiveness of the proposed approach.

3 - Tackling Transfer Learning Problems Optimally: Methods for Regression Under Covariate Shift

Liviu Aolaritei, UC Berkeley, Berkeley, CA, United States, Michael I. Jordan, Reese Pathak, Annie Ulichney

Modern machine learning methods are often deployed in scenarios where the training and test data distributions differ. Motivated by these distribution shifts, in this work, we study linear regression under a form of distribution shift known as covariate shift. When the covariance matrix of the training distribution is appropriately bounded by the covariance matrix of test distribution, we derive computationally efficient and statistically optimal estimators. Our work quantifies how the interaction between constraints on the underlying regressors and the level of discrepancy between the training and test distributions determine the statistical complexity of such learning problems.

4 - Learning in a Changing World: Perspectives from Abstention and Relaxed Benchmarks

Abhishek Shetty, Massachusetts Institute of Technology, Cambridge, MA, United States

Though modern machine learning has been highly successful, as we move towards more critical applications, many challenges towards building trustworthy systems, such as ensuring robustness, privacy, and fairness, arise. Ad hoc and empirical approaches have often led to unintended consequences for these objectives, thus necessitating a principled approach. Traditional solutions often require redesigning entire pipelines, come with a significant loss in quality or typically make strong assumptions on the data. In this talk, we will look at principles towards incorporating important desiderata into existing pipelines without significant computational and statistical overhead. Towards addressing this, we look at two perspectives on prediction on arbitrary data sequences: abstention and relaxed benchmarks. We will show that using these perspectives allows us to circumvent impossibility results regarding prediction. Further, we will convey connection between these perspectives and other perspectives on beyond worst case analysis of learning. The talk will be based on joint works with Surbhi Goel, Steve Hanneke, Shay Moran, Omar Montasser and Nikita Zhivotovsky.

5 - Improved Rates for Online Calibration

Noah Golowich, MIT, Cambridge, MA, United States, Yuval Dagan, Constantinos Daskalakis, Maxwell Fishelson, Robert Kleinberg, Mehryar Mohri, Princewill Okoroafor, Jon Schneider

A set of probabilistic forecasts is calibrated if each prediction of the forecaster closely approximates the empirical distribution of outcomes on the subset of timesteps where that prediction was made. In this talk we discuss the fundamental problem of online calibrated forecasting, in which an adversary chooses the outcomes which are revealed to the forecaster in a sequential manner.

We begin by studying the setting where the predictions are binary, which was initiated by Foster & Vohra (1998). They derived an algorithm with $T^{2/3}$ calibration error after T time steps, and showed a lower bound of $T^{1/2}$. We give the first improvement to the $T^{2/3}$ upper bound on calibration error of Foster & Vohra. To do so, we study a combinatorial game called sign preservation recently introduced by Qiao & Valiant (2021) and show new upper bounds on the value of this game.

In the second half of the talk, we move to the problem of high-dimensional calibration, where the forecasts belong to an arbitrary convex subset of Euclidean space. We connect this problem with that of regret minimization for online linear optimization, showing that upper bounds for regret minimization imply the existence of efficient calibrated algorithms. In the special case that the convex subset is the simplex, our result recovers a recent result of Peng (2025). Moreover, we establish a lower bound showing that our result is near-optimal, significantly improving a corresponding lower bound of Peng for this problem.

TD21

Building A Level 4 A407

Electrification of Urban Transportation

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Dongyang Xia, Delft University of Technology, Delft, N/A, Netherlands

1 - A Bilevel Optimization for Electric Vehicles Charging Stations Placement

Mobina Nankali, University of Minnesota, Minneapolis, MN, United States, Michael Levin

Planning electric vehicle (EV) charging infrastructure requires balancing economic goals with user convenience and demand patterns. A bilevel optimization framework is developed to capture the strategic decisions of network planners, who aim to maximize revenue through optimal station placement, and the responsive behavior of EV users, who seek to minimize travel time and charging costs. A solution method is proposed, combining High Point Relaxation, Value Function Cuts, and Outer Approximations to refine the solution space systematically. A Branch-and-Price-and-Cut (BPC) algorithm further enhances computational efficiency by integrating column generation, allowing the model to scale with large user path sets. Results demonstrate that explicitly modeling user behavior leads to more efficient and profitable infrastructure outcomes, improving station utilization and reducing investment waste.

2 - First-in-first-out (FIFO) or highest-energy-first-out (HEFO)? A tale of two electric vehicle (EV) battery swapping policies

Mengyi Sha, Tsinghua University, Beijing, China, People's Republic of

3 - Data-Driven Battery Swapping Routing Methods for E-Bike Sharing Systems

Xinyu Jiang, Tsinghua University, Beijing, China, People's Republic of, Wei Qi

Shared e-bike systems have emerged as a vital component of sustainable urban mobility. Efficiently managing their battery swapping operations, specifically the routing of service staff, is paramount for ensuring e-bike availability and overall system viability. However, current approaches often optimize limited short-term metrics and compromise long-term value. They struggle with dynamic adaptation, fairness integration and scalability due to inflexible parameter setting. This paper presents a novel data-driven framework to address these challenges through the integration of continuous-time Markov chains (CTMCs) and reinforcement learning (RL). A CTMC is proposed to model the state of e-bikes in each sub-region. Based on the states of sub-regions, a spatio-temporal network model is established to capture the decision process of battery swapping routing, with the objective of maximizing the long-term value of regional operations. Adaptability and human-centricity are enhanced via a feedback mechanism that leverages large language models to interpret staff input. The optimization of routing model parameters is achieved through an RL method guided by a comprehensive reward function which incorporates efficiency, fairness and service metrics. Collaborating with one of the leading shared e-bike operators in China, we validate the proposed method using both synthetic data and test instances built on their real operational dataset. Extensive experiments across multiple scenarios demonstrate that our approach outperforms existing strategies in terms of computational efficiency, battery swapping performance, fairness and scalability. This research contributes an advanced artificial intelligence-assisted methodology for battery swapping routing and offers practitioners an adaptive solution to improve operational performance in shared mobility systems.

4 - Robust Charging Station Location and Routing-Scheduling for Electric Modular Autonomous Units

Dongyang Xia, Delft University of Technology, Delft, Netherlands, Lixing Yang, Yahan Lu, Shadi Sharif Azadeh

Motivated by global electrification targets and the advent of electric modular autonomous units (E-MAUs), this paper addresses a robust charging station location and routing-scheduling problem (E-RCRSP) in an inter-modal transit system, presenting a novel solution to traditional electric bus scheduling. The system integrates regular bus services, offering full-line or sectional coverage, and short-turning services. Considering the fast-charging technology with quick top-ups, we jointly optimize charging station locations and capacities, fleet sizing, as well as routing-scheduling for E-MAUs under demand uncertainty. E-MAUs can couple flexibly at different locations, and their routing-scheduling decisions include sequences of services, as well as charging times and locations. The E-RCRSP is formulated as a path-based robust optimization model, incorporating the polyhedral uncertainty set. We develop a double-decomposition algorithm that combines column-and-constraint generation and column generation armed with a tailored label-correcting approach. To improve computational efficiency and scalability, we propose a novel method that introduces super travel arcs and network downsizing methodologies. Computational results from real-life instances, based on operational data of advanced NExT E-MAUs with cutting-edge batteries provided by our industry partner, indicate that charging at both depots and en-route fast-charging stations is necessary during operations. Moreover, our algorithm effectively scales to large-scale operational cases involving entire-day operations, significantly outperforming state-of-the-art methods. Comparisons with fixed-composition buses under the same fleet investment suggest that our methods are able to achieve substantial reductions in passengers' costs by flexibly scheduling units. Our approach is appealing to operators, providing entire-day solutions with enhanced service quality.

5 - National Electric Truck Charging Network Planning with Cluster-Based Optimization and Temperature Sensitivity

Hanieh Rastegar Moghaddam Bajestani, University of Illinois at Chicago, Chicago, IL, United States, Xi Cheng, Xueqi Zeng, Amir Shafiee, Jialu Yang, Chi Xie, Jane Lin

This study develops a scalable optimization framework for electric truck (ET) charging infrastructure planning, explicitly incorporating temperature-dependent battery efficiency and network-wide flow patterns across the United States. Unlike previous approaches, which address the problem on a national scale as a monolithic network, we propose a novel hierarchical decomposition approach that enables computational tractability over the U.S. national interstate highway network. The proposed method first applies spatial clustering to partition the national highway network into several geographically coherent regions based on freight flow density and spatial proximity derived from and Federal Highway Administration (FHWA)'s Freight Analysis Framework (FAF) data. Within each cluster, a bi-level optimization model is solved to jointly optimize charging station location, sizing, and long-haul ET routing, while capturing user equilibrium under local temperature variations and queuing delays using a convex combinations algorithm embedded with a multi-criterion label-correcting method. To address inter-cluster flows, a corridor-based routing strategy is proposed to optimize key long-haul connections between clusters. These sub-network solutions are then integrated to form a comprehensive charging infrastructure plan that balances intra-regional efficiency and national connectivity. By decomposing the full-scale problem and capturing both local operational constraints and cross-regional interactions, our approach provides a practical and flexible planning framework for nationwide ET deployment. The updated model will be validated with the FAF commodity and truck flows. This research offers policy insights into regionalized infrastructure investment, adaptive planning under environmental constraints, and scalable frameworks for decarbonizing freight transportation at the national level.

TD22

Building A Level 4 A408

Processes and Algorithms on Networks

Invited Session

Applied Probability Society

Chair: Souvik Dhara, Purdue University, West Lafayette, IN, United States

1 - Testing a Planted Matching in a Random Graph at the Critical Threshold

Cheng Mao, Georgia Institute of Technology, Atlanta, GA, United States

To study inference of hidden information in statistical networks, researchers have been working extensively on random graph models with planted structures in recent years. In this work, we focus on testing the presence of a planted matching in a random graph. Our results not only identify the statistical threshold for this problem but also establish the asymptotic normality of the log-likelihood ratio at the critical threshold. To achieve this, we provide two approaches, one based on an orthogonal decomposition of the likelihood ratio and the other based on the cluster expansion of the log-likelihood ratio. Either approach shows that the likelihood is dominated by the signed wedge count, a simple statistic that is both statistically optimal and computationally efficient.

2 - Propagation of Shocks on Networks: Can Local Information Predict Survival?

Souvik Dhara, Georgia Institute of Technology, Atlanta, GA, United States

Complex systems are often fragile, where minor disruptions can cascade into dramatic collapses. Epidemics serve as a prime example of this phenomenon, while the 2008 financial crisis highlights how a domino effect, originating from the small subprime mortgage sector, can trigger global repercussions. Similarly, a massive power outage in eastern North America was seemingly set off by a localized fault. The mathematical theory underlying these phenomena is both elegant and foundational, profoundly shaping the field of Network Science since its inception. In this talk, I will present a unifying mathematical model for network fragility and cascading dynamics, and explore its deep connections to the theory of local-weak convergence, pioneered by Benjamini-Schramm and Aldous-Steele.

3 - Mixing Phases and Metastability for the Glauber Dynamics on the P-Spin Curie-Weiss Model

Somabha Mukherjee, National University of Singapore, Singapore, Singapore, Ramkrishna Samanta, Jiang Zhang

In this talk, we will analyze the mixing behavior of Glauber dynamics for the p-spin Curie-Weiss model with an external field. By examining the structure of a related negative free-energy function, we will identify three disjoint regions in the parameter space, each corresponding to a different order of mixing time. These regimes depend on the number and nature of local maxima: fast mixing occurs when there's a unique, strongly concave maximum; exponentially slow mixing arises in the presence of multiple local maxima (metastability); and a critical regime with a flat maximum leads to polynomially slower mixing. We give a near-complete geometric characterization of these regions and show that the remaining portion of the space—a one-dimensional boundary curve—corresponds to an unresolved case involving

stationary inflection points, where one can establish a polynomial lower bound for the mixing time. Finally, we will demonstrate that even in the metastable regime, one can define a restricted version of the original Glauber dynamics that mixes rapidly within a single local basin.

4 - The DeGroot Model on Complex Networks

Fuwei Yu, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

The DeGroot model is a popular model for opinion formation on directed networks, where at each time-step individuals update their opinions by taking a weighted average of their neighbors' opinions, either synchronously or asynchronously. Provided the matrix of weights is irreducible, the model under synchronous updates is known to converge to the Perron-Frobenius eigenvector times the vector of initial opinions, resulting in consensus. In this talk we study the limiting opinions in the DeGroot model when the underlying network is generated through a random graph model having as a local weak limit a Galton-Watson process (single or multitype) and characterize the limit in terms of an associated weighted branching process.

TD23

Building A Level 4 A409

Advanced Analytics

Contributed Session

Contributed

Chair: Lingqing Shen, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Optimal Immutable Series Selection in Forecast Reconciliation for Hierarchical TIME Series

Wai Yin Leung, Monash University, Melbourne, Australia, Shanika Wickramasuriya

Hierarchical time series (HTS) represent collections of time series organized according to aggregation constraints, such as geographical or product-based hierarchies. Forecasts for individual series in an HTS are typically generated independently, resulting in incoherence. Forecast reconciliation refers to the process of adjusting these base forecasts to ensure coherence. In standard reconciliation approaches, all forecasts are adjusted. However, in many practical settings, it may be necessary—or even beneficial—to keep certain forecasts unchanged. These are referred to as *immutable series*. Currently, the choice of which series to designate as immutable is often heuristic or based on domain expertise. This paper proposes a novel mathematical formulation and an iterative cutting plane algorithm to optimise the selection of immutable series in forecast reconciliation. We assess the scalability of the method via simulation and demonstrate its superior performance using Australian domestic tourism data.

2 - Ad Assignment with Social Influence: A Scalable Optimization Approach

Reyhaneh Kharazmi Pour, University of California, Irvine, Irvine, CA, United States, John Turner

In digital advertising, some advertisers prefer to show their ads to viewers who are socially connected, such as friends. The idea is that when friends are exposed to the same ad, it can increase influence and lead to higher engagement and conversion. Allowing for this kind of targeting creates new opportunities for publishers to generate more revenue, but it also makes the ad allocation problem more complex. We study an ad delivery setting where the value of assigning an ad to a viewer depends not only on the individual but also on whether their friends are assigned the same ad. This creates a need to jointly consider individual impact, social influence, advertiser budgets, and constraints across the entire viewer network. To address this, we develop an efficient algorithm that produces high-quality ad assignments while accounting for social connections. The algorithm is designed to scale to large instances and to find near-optimal solutions without requiring excessive computational resources.

3 - Optimal Forecast Reconciliation for Quantiles

Anastasios Panagiotelis, Monash University, CAULFIELD EAST, Australia, Hossein Alipour, George Athanasopoulos, Nam Ho-Nguyen

A new methodology is developed for reconciling probabilistic forecasts where the objective is to obtain accurate forecasts for a given quantile level. This is achieved by minimizing expected pinball loss, a challenging problem which we propose to overcome in two ways. First, to allow for full generality of base forecasts, expectations are approximated by drawing a sample from the base forecast. Second, to allow for optimization via first order methods, pinball loss is approximated by a smooth function. Theoretical results are developed proving that the minimizer of the objective function that employs these two approximations converges to the minimizer of expected pinball loss in the limit. The newly developed methodology is then applied in a simulated and real-world empirical setting and shown to significantly outperform the very competitive MinT benchmark.

4 - Dynamic Pricing Strategy for Perishable Products Considering Price and Age-Dependent Logit Demand Function

Moddassir Khan Nayeem, University of Texas at San Antonio-Department of Mechanical, Aerospace, and Industrial Engineering, San Antonio, TX, United States, Hamid Najafzad, Fuhad Ahmed Opu, Yara Altarawneh, Omar Abbaas

Effective pricing strategies are critical for managing perishable products due to their limited shelf life and declining consumer appeal as they age. This study introduces a comprehensive framework for dynamic pricing optimization, combining an exact algorithm, inflection point-based analysis, and six discount policies. The exact algorithm identifies revenue-maximizing prices for each product age by solving a logistic demand function using the Lambert W function, providing a precise solution. Building on this, the inflection point-based algorithm pinpoints ages where consumer sensitivity to price is highest, guiding price adjustments that sustain demand and enhance revenue. Six discount policies, including uniform, age-specific, threshold-driven, and lifecycle-based approaches, are proposed to address varying market dynamics. These policies target specific stages of the product lifecycle to balance inventory turnover and profitability. Results demonstrate significant demand and revenue improvements compared to baseline inflection point-based pricing. Uniform discounts at inflection points maximize total revenue, while lifecycle-based strategies achieve balanced performance across all ages. Policies tailored to specific stages, such as threshold-driven and uniform discounts for younger or older products, excel in inventory clearance or early profitability. This research offers actionable insights into aligning pricing strategies with consumer behavior and product lifecycle dynamics, equipping retailers with data-driven tools to optimize revenue, reduce waste, and enhance inventory turnover for perishable products.

5 - Auto-Calibration and Biconvex Compressive Sensing: Theory and Applications to Parallel MRI

Yuan Ni, University of California Davis, Davis, CA, United States, Thomas Strohmer

We study a class of blind inverse problems in which a transform-sparse signal is measured by multiple sensors with unknown system parameters. This setup naturally arises in a range of applications, including imaging, wireless sensing, and biomedical signal acquisition, where explicit calibration is difficult or impractical. The problem leads to a biconvex formulation involving simultaneous recovery of the signal and unknown sensing operators — a setting that is typically challenging due to the lack of theoretical recovery guarantees.

We address this challenge by reformulating the problem as a convex optimization task using the 'lifting' technique. By leveraging the signal's sparsity structure and the redundancy introduced by multiple sensors, we solve a mixed-norm minimization problem to jointly recover the signal and sensing parameters. We derive robust and stable recovery guarantees under noise and sparse deficiencies.

As a key application, we demonstrate the approach on self-calibrated parallel MRI (pMRI), where explicit coil sensitivities are often unavailable. Our method enables auto-calibrated imaging with theoretical guarantees and closes a longstanding theoretical gap, aligning with empirical findings reported in the pMRI literature.

6 - Parametric Workload Distribution for Balanced Optimization: Theory and Application

Abhay Singh Bhadoriya, Los Alamos National Laboratory, Los Alamos, NM, United States

Workload distribution is a fundamental challenge across various operations research applications, including multi-agent path planning, task assignment, and facility location. In general, workload balancing in all of these applications can equivalently be formulated as a problem of minimizing the dispersion of a vector of decision variables. In this talk, we will present a general parametrized family of constraints to control the dispersion of a vector of decision variables. We introduce a parameter to control the amount of dispersion and show that, in certain special cases, these constraints reduce to a single second-order conic constraint or a set of linear constraints. We will also discuss some theoretical properties for the same. Finally, we demonstrate the efficacy of the proposed formulations using off-the-shelf or existing methods on a few practical examples from vehicle routing, multi-agent task allocation and power grids.

7 - Linx Double-Scaling: An Enhanced Relaxation with Efficient Algorithm for the Maximum-Entropy Sampling Problem

Lingqing Shen, Carnegie Mellon University, Pittsburgh, PA, United States, Fatma Kılınc-Karzan

In this work, we study convex relaxations for the Maximum-Entropy Sampling Problem (MESP), a key problem in experimental design. We first use majorization theory to derive convex relaxations for a general class of principal submatrix selection problems, which recover the state-of-the-art DDFact [Li and Xie, 2024] and linx [Anstreicher, 2020] relaxations. We further analyze the majorization-based relaxations by establishing convex hull properties and equivalent representations. Next, we introduce a new double-scaling technique for the linx relaxation and prove that it admits a convex-concave saddle point formulation. In addition, we address two open questions posed by Chen et al. [2024], showing the convex-concavity of g-scaling for the DDFact relaxation and proving its equivalence to the original DDFact relaxation. Building on these results, we reveal a new connection between the DDFact and linx relaxations, both in their original and scaled forms, which, in particular, yields the first known bound on the integrality gap for linx. Finally, to efficiently solve the scaling relaxations as saddle point problems, we propose parameter-free variants of the extragradient algorithm with adaptive stepsize schemes and prove a non-ergodic convergence rate of $O(1/\sqrt{T})$. Our experimental results highlight the improved quality of the proposed linx double-scaling compared to g-scaling and o-scaling from the literature, and demonstrate the practical efficacy of our algorithm.

TD24

Building A Level 4 A410

Statistical and Algorithmic Foundations of Reinforcement Learning

Tutorial session

Tutorials

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

1 - Statistical and Algorithmic Foundations of Reinforcement Learning

Yuxin Chen, University of Pennsylvania, Philadelphia, PA, United States, Yuejie chi, Yuting Wei

As a paradigm for sequential decision making in unknown environments, reinforcement learning (RL) has received a flurry of attention in recent years. However, the explosion of model complexity in emerging applications and the presence of nonconvexity exacerbate the challenge of achieving efficient RL in sample-starved situations, where data collection is expensive, time-consuming, or even high-stakes (e.g., in clinical trials, autonomous systems, and online advertising). How to understand and enhance the sample and computational efficacies of RL algorithms is thus of great interest. In this tutorial, we aim to introduce several important algorithmic and theoretical developments in RL, highlighting the connections between new ideas and classical topics. Employing Markov Decision Processes as the central mathematical model, we cover several distinctive RL scenarios (i.e., RL with a simulator, online RL, offline RL, robust RL, and RL with human feedback), and present several mainstream RL approaches (i.e., model-based approach, value-based approach, and policy optimization). Our discussions gravitate around the issues of sample complexity, computational efficiency, as well as algorithm-dependent and information-theoretic lower bounds from a nonasymptotic viewpoint.

TD25

Building A Level 4 A411

Network Design

Invited Session

TSL: Freight Transportation

Chair: Christian Truden, University of Klagenfurt, Klagenfurt, 9020, Austria

1 - Cost-Effective and Speedy Fulfillment for a Large Assortment to a Large Number of Customers

Kenneth Kuhn, Walmart, Seattle, WA, United States

Walmart would like to reduce fulfillment costs and increase delivery speeds for a large and increasing assortment of items to hundreds of millions of customers in the United States in the next five years. To accomplish this, we first forecast future demand, in time and space and by item group. We identify clusters of dense demand. Such areas can be cost-effectively served using delivery vans from a nearby Walmart facility. We go on to find optimal sets of Fulfillment Centers, Sortation Centers, and Delivery Stations to serve the demand clusters. We solve a four-layer transshipment optimization problem to produce an optimal fulfillment policy. Finally, we solve prototype vehicle routing problems to construct models of Delivery Station operations and analyze unit economics. We find that no single-layer Fulfillment Network or single delivery method or mode is capable of achieving our desired results. The lowest per package costs are achieved by using relatively full delivery vans in areas of dense demand. It makes sense for these vans to pick up packages at Sortation Centers and Delivery Stations that collect shipments from multiple Fulfillment Centers. Third-party express delivery services are particularly useful for last-mile deliveries in areas with sparse demand. We set out to identify the optimal, near-future, multi-layer Walmart Fulfillment Network and network flows of goods within this network and to customers. We needed to develop formulations of network design and network flow problems tailored to the Walmart use case. We needed to define solution search strategies that scale well.

2 - Hydrogen Refueling Station Location Problem Considering Pipeline Delivery

Jane Lin, University of Illinois Chicago, Chicago, IL, United States, Jialu Yang, Bo Zou

Hydrogen Fuel Cell Trucks (HFCTs) are promising alternatives to Diesel Trucks (DTs) for long-haul freight transportation in reducing greenhouse gas emissions while maintaining similar operational characteristics to DTs. Despite continued advancement in hydrogen technologies, investigation into strategic deployment of Hydrogen Refueling Stations (HRS) along the U.S. interstate highway network remains limited. Thus, this study considers an optimal hydrogen refueling station network design problem that minimizes the total cost of HRS construction and hydrogen delivery. In particular, we consider a hybrid hydrogen delivery method that utilizes the existing natural gas pipeline network throughout the country to deliver hydrogen from the production hubs to the storage facilities (reservoirs) and then trucks to delivery hydrogen from reservoirs to HRS. The problem is formulated as a Mixed-Integer Linear Program (MILP) that satisfies interstate highway freight demand using the Freight Analysis Framework (FAF) for 2050. A Lagrangian Relaxation-based decomposition algorithm is proposed and its performance is evaluated against Gurobi in terms of solution quality and computational efficiency. Our results highlight the predominance of HRS construction costs, constituting over 70% of total cost. Furthermore, pipeline-based hydrogen transport exhibits outstanding economic advantages over truck-based hydrogen delivery, underscoring the strategic importance of leveraging existing pipeline infrastructure and advocating for a dedicated hydrogen pipeline network in the future. The findings provide valuable policy implications for targeted governmental support in hydrogen distribution and refueling infrastructure building to facilitate economically viable adoption of HFCTs for long haul freight transportation.

3 - A Continuous Approximation Approach for Designing Collection and Delivery Point Networks

Benjamin Rojas, Pontificia Universidad Católica de Chile, Santiago, Chile, Homero Larrain, Mathias Klapp

Collection and delivery points (CDPs) are facilities where customers can pick up their online purchases, enabling logistics operators to consolidate multiple deliveries into a single stop and reduce distribution costs. However, the choice of the customer depends on factors such as personal preferences, proximity to the CDP network, and

economic incentives offered by logistics providers in the form of discounts or cashback. In this study, we propose a strategic model for designing a cost-efficient next-day last-mile distribution system that incorporates CDP pickups, balancing infrastructure investments, transportation costs, and customer incentives. Using a continuous approximation approach, we estimate total costs and determine the optimal number of CDPs and incentive levels. Key findings highlight that optimal CDP installation significantly reduces costs, especially in high-density demand areas. Additionally, incentives are especially effective in low-density areas, encouraging customers to travel further and reducing the need for additional CDPs. Our experiments also suggest that CDPs, without incentives, can reduce costs by 16.9%, while adding incentives increases savings to 28.1%. This research offers a comprehensive framework for understanding and designing cost-efficient urban parcel distribution systems.

4 - A Framework for Measuring Maritime Port System Capacities with Limited Input Data

Debojjal Bagchi, The University of Texas at Austin, Austin, TX, United States, Kyle Bathgate, Stephen Boyles

We present a framework to estimate the “operating capacity” and “ultimate capacity” of a multimodal maritime freight system. Accurately determining the capacity of a port aids stakeholders in making informed decisions about infrastructure investment and resource allocation. Traditional methods for port capacity assessment often rely on data-driven simulations, which can be difficult to calibrate due to complex interactions among port systems and the large volume of data required. We present a data-efficient alternative to estimate port capacities by demonstrating that anchorage queueing behavior and channel entry patterns serve as effective indicators of port performance. We develop a novel queueing-theory-based formulation to define the operating capacity of a port system as the maximum vessel arrivals that can be sustained for an extended period of stable operations. To capture the port’s performance beyond stable operations, we propose a dynamic model to estimate the ultimate capacity, or the maximum throughput that the port can handle in saturated conditions. Our approach requires minimal data, which can be readily obtained from archival automatic identification system (AIS) vessel trajectories, making implementation robust and efficient. We validate our framework using a discrete-event simulation of the Port of Houston and archival AIS data. Our model estimates that the port’s operating capacity is about 5% above current levels, while ultimate capacity is approximately twice that of the operating capacity. Anchorage wait times predicted by our model closely match historic vessel tracking data with errors in the range of 0.35% to 17%, indicating the viability of the proposed model.

5 - A Heuristic Approach for the Service Network Design with Fleet and Emissions Management

Christian Truden, University of Klagenfurt, Klagenfurt, Austria, Mike Hewitt

Less-than-truckload (LTL) carriers often operate transportation networks covering vast geographic areas - frequently spanning the entire contiguous United States - and deploy large fleets of trucks. In these networks, resource prices vary significantly from region to region, especially between states as well as between urban and rural areas. The complexity increases further when carriers consider introducing sustainable power-train technologies such as battery-electric or hydrogen-electric heavy-duty tractors. The Service Network Design problem with Fleet and Emissions Management (SND-FEM) accounts for these factors by including multiple power-train technologies and incorporating spatial variations in fuel prices and Scope 2 emissions associated with the energy sources used by these tractors. However, LTL carriers face very large SND-FEM instances, making exact solution methods inapplicable. To address this challenge, we propose a path-based heuristic that constructs and explores alternative routes between network terminals. Our approach employs a two-level representation of the underlying road network, which simplifies resource-related considerations during path construction. We present initial results demonstrating the effectiveness of our heuristic in solving realistic SND-FEM scenarios.

TD26

Building A Level 4 A412

Metamodeling and Design of Experiments

Invited Session

Simulation

Chair: Wei Xie, Northeastern University, Boston, MA, United States

Co-Chair: Qiong Zhang, Clemson University, Clemson, SC, United States

1 - Space-Filling MOFAT Designs

Roshan Joseph, ISyE Georgia Tech, Atlanta, GA, United States, Difan Song

Space-filling designs are commonly used for metamodeling. They ensure that the testing points are not too far from the training points and thus provide a robust prediction performance. However, they cannot quickly identify the important factors from the set of factors used in the simulation. Screening designs such as the maximum one-factor-at-a-time (MOFAT) designs do a much better job for identifying the important factors. This talk will explain an improvement to MOFAT designs that aims at both prediction and screening.

2 - Enhancing Underutilized Bus Routes with Semiflexible Routing

Ye Chen, Bowling Green State University, Bowling Green, OH, United States, Md Hishamur Rahman, Shijie Chen, Yanshuo Sun, Muhammad Siddiqui, Matthew Mohebbi, Nikola Markovic

We explore improving an underutilized conventional bus route by converting it into a semiflexible transit system where passengers provide advance notice of their intended stops, allowing buses to skip downstream stops without demand by taking shortcuts. This approach increases stop density, reduces walking distances to and from bus stops, and maintains operational efficiency. To design this system, we develop optimization models that maximize the number of stops while adhering to tour duration and arrival time constraints. Then, a Bayesian framework is established to analyze the sequential simulation experiment outputs, so that we can (i) reduce the number of simulation experiments needed for assessing violation of constraints, and (ii) decrease the duration of individual simulation runs by aborting them immediately once the bus tour duration surpasses the desired upper bound.

3 - Sequential Experimental Design for Adaptive Reduced-Order Modeling

Jingyi Feng, Georgia Institute of Technology, Atlanta, GA, United States, Xiao Liu

Projection-based Reduced-Order Modeling requires generating data from full-order systems (through numerical methods). Because a full-order system is often parameterized by a set of parameters, challenges immediately arise when one needs to (sequentially) determine the parameter settings, from the parameter space, under which the full-order system is solved and data are generated. In this talk, we present some preliminary results on the sequential experimental design for constructing reduced-order models that well capture the system behavior over the parameter space (needed in design, control, uncertainty quantification and real-time operations). This talk consists of two major parts. In the first part, we describe a Projected Gaussian Process (pGP) that establish the relationship between parameters and reduced-order models. In the second part, we share some early results on a sequential design framework that enables us to sequentially construct or update the pGP model by collecting data from the optimal experimental conditions (i.e., parameter settings).

4 - Bioprocess Mechanism Bayesian Learning and Monitoring

Wei Xie, Northeastern University, Boston, MA, United States

While biomanufacturing plays an increasingly significant role in supporting public health and fighting pandemics, it faces critical challenges including high complexity, high variability, and very limited historical process experiment runs. Driven by these challenges, we propose a novel Bayesian learning framework on bioprocess multi-scale mechanistic and hybrid model in the form of stochastic differential equations (SDEs). The proposed framework can integrate heterogeneous and sparse data, facilitate the learning of underlying regulatory mechanisms and support monitoring latent state trajectories.

5 - Surrogate-Based Contextual Simulation Optimization: A Divide-and-Conquer Approach with Application to Distribution Inventory Systems

Junkai Zhao, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Weiwei Fan, Jun Luo, Tong Wang, Xiaowei Zhang

Distribution inventory system management plays a pivotal role in modern retail supply chains, directly influencing operational efficiency and market responsiveness. However, these systems confront critical challenges stemming from inherent high stochasticity, high context-dependency, and computational intensity. To address these limitations, we propose a contextual simulation optimization framework employing surrogate modeling. Our methodology utilizes Gaussian processes (GPs) to construct a metamodel that captures the joint dependence of cost functions on inventory policies and contextual variables, enabling data-driven policy optimization. To resolve computational bottlenecks associated with inverting large covariance matrices in conventional GP implementations, we develop a parallelizable divide-and-conquer

algorithm. This approach strategically partitions datasets into subsets for localized GP training, subsequently aggregating predictions through ensemble averaging. We derive asymptotic bounds for convergence rate of the optimality gap in our algorithm. Extensive numerical experiments validate the ability of our method to handle industrial-scale inventory problems and demonstrate its consistent effectiveness and efficiency across various inventory networks.

TD27

Building B Level 2 B201

Software for Optimization Modeling and Deployment 1

Invited Session

Computing Society

Chair: Susanne Heipcke, FICO, Xpress Optimization, Birmingham, United Kingdom

Co-Chair: Robert Fourer, AMPL Optimization Inc., Evanston, IL, United States

1 - Recent New Features of Fico Xpress APIs and Xpress Mosel

Susanne Heipcke, FICO, Xpress Optimization, Birmingham, United Kingdom

This talk gives an overview of recent updates to the FICO Xpress APIs. Primary focus will be on the Python API for Xpress that has been reworked to align it more closely with the other APIs (see <https://github.com/fico-xpress/python-notebooks>), and new features of Xpress Mosel, including its tools and components (see <https://github.com/fico-xpress/mosel>).

We also shortly outline updates to the object-oriented APIs (C++, C#, Java) and to the recently published Julia API for Xpress that is provided open-source on Github (see <https://github.com/fico-xpress/XpressAPI.jl>) and through the official Julia registry.

2 - Efficient Data Exchange in Amplpy via Apache Arrow Integration

Jurgen Lentz, AMPL Optimization Inc., Mountain View, CA, United States

AMPLPy is the most popular Python API for AMPL, a powerful algebraic modeling language for large-scale mathematical optimization. As modern optimization workflows increasingly integrate with data-intensive environments, efficient data handling becomes critical. To address this, AMPLPy now incorporates support for Apache Arrow, an open-source, columnar in-memory data format designed for high-performance analytics.

By leveraging Apache Arrow, AMPLPy enables rapid, zero-copy data transfer between AMPL models and Python data structures such as Pandas, Polars and many more data frames. This integration minimizes data transfer overhead and significantly improves the speed and scalability of data exchange in optimization pipelines. Users can seamlessly pass large datasets from Python into AMPL for model instantiation and retrieve results back into Python with minimal latency. Arrow's cross-language and cross-platform capabilities enhance interoperability with other systems and data science tools, and will also be used to transfer data to and from other AMPL APIs, such as those in Go and Rust, in the future.

The use of Arrow aligns AMPLPy with modern data engineering practices, ensuring compatibility with distributed data processing frameworks and in-memory analytics platforms. This development makes AMPLPy an even more attractive tool for both researchers and practitioners looking to embed optimization within data-centric and performance-critical applications.

3 - Beyond the Solver: Designing Uis that Empower Decision-Makers

James Laramore, FICO, Hillsborough, NC, United States

Large-scale optimization models generate tremendous value, but their real-world impact often hinges on how easily end-users can interact with, deploy, and act on the model's results. This session explores how thoughtfully designed user interfaces (UIs) bridge the gap between optimization experts and decision-makers, enabling faster, more confident operational decisions.

We'll highlight best practices for creating model-driven applications and show how integrating generative AI and large language models (LLMs) enhances user engagement through practical, natural language explanations and interactive decision support. Attendees will learn strategies for building trust, accelerating adoption, and decreasing time-to-value through interactive, decision-ready user interfaces.

4 - Exploring the Interplay Between LLMs and Mathematical Optimization Model Building

Erik Halseth, Gurobi Optimization, Los Angeles, CA, United States

Generative AI based tools are being positioned to transform the workflows surrounding mathematical optimization. At Gurobi, we want to leverage new opportunities in this space while still honoring the importance of studying and practicing classical model development techniques. In this presentation, we'll walk you through some of our recent work aimed at equipping users with knowledge and tools for leveraging LLMs to develop and deploy optimization models with gurobipy. We will hone in a hypothetical workflow that will serve to highlight key aspects on bridging together LLMs and mathematical optimization models. Whether you're starting your optimization journey or fine-tuning your skills, discover how these innovative AI resources can streamline your workflow and unlock new levels of efficiency.

TD28

Building B Level 2 B202

Advances in Multimodal Learning for Healthcare

Invited Session

Computing Society

Chair: Matthew Peroni, MIT, Cambridge, MA, United States

1 - When is More Better? Efficient Personalized Modality Selection

Matthew Peroni, MIT, Cambridge, MA, United States

The ultimate goal of multimodal learning is to combine disparate and complementary information from multiple data modalities, such as images, text, sensors, and audio, to create more robust, generalizable, and accurate models. However, incorporating many modalities can increase costs and introduce additional complexity into deployed ML systems. Previous work has investigated modality selection at the population level, selecting a subset of modalities that provide the best balance of performance and costs. In this work, we go beyond the population level, introducing the idea of personalized modality selection, which adaptively selects modalities to acquire based on some contextual information about the sample. We characterize this problem setting and its challenges and introduce efficient algorithms for scaling personalized modality selection to scenarios where the total number of modalities is very large. Applying this approach to clinical decision support tasks, we achieve improved performance over baselines while reducing the total cost of data acquisition substantially.

2 - Multimodal Machine Learning for Clinical Trajectory Prediction in Psychiatric Care

Jiayi Gu, Massachusetts Institute of Technology, Cambridge, MA, United States, Dimitris Bertsimas

Predicting prognostic outcomes—such as length of stay, relapse, and risk of admission—remains a critical challenge in clinical psychiatric care. Many psychiatric conditions, including psychosis, mood disorders, and anxiety disorders, are characterized by fluctuating courses, where relapse is often preceded by subtle but detectable symptom changes weeks or even months in advance. However, current methods have limited predictive power and often fail to support timely intervention. This study leverages multimodal electronic medical record (EMR) data to develop machine learning algorithms that predict clinical trajectories and outcomes across a range of psychiatric diagnoses. By integrating diverse data modalities—including structured clinical data, unstructured clinical notes, and treatment histories—our approach aims to improve prognostic accuracy. The ultimate goal is to enhance patient outcomes, reduce the economic burden on health systems, and demonstrate the value of predictive models in enabling proactive, personalized mental healthcare.

3 - Multi-Class Diagnosis of Left Ventricular Systolic Dysfunction from Electrocardiograms

Catherine Ning, MIT, Cambridge, MA, United States, Dimitris Bertsimas, Yu Ma, Cindy Wang

A reduced left ventricular ejection fraction (LVEF) serves as an essential diagnostic and prognostic marker for heart failure and risk of sudden death. Traditionally, the estimation of ejection fraction has relied on time-consuming and expensive imaging techniques. In contrast, the standard 12-lead electrocardiogram is much faster, more affordable, and is the most widely used complementary exam for heart evaluation, yet does not allow for a direct human interpretation of the LVEF. Therefore, we believe that an accurate AI-enabled diagnosis system is most acutely needed and can be further improved by analyzing large amounts of available patient information. This can be of immense value especially in outpatient or resource-limited settings, e.g., in low- and middle-income countries, where individuals often do not have access to trained cardiologists.

Most AI models for cardiac diagnosis proposed in the literature have focused on unimodal, binary classification from ECGs, often limited to identifying moderately or severely reduced LVEF while unable to capture mildly reduced LVEF. In contrast, we predict more granular levels of LVEF ranges, specifically, normal, mildly, moderately, or severely reduced ejection fraction. Additionally, we leverage both ECG tracings and patient electronic health records, thereby offering a cost-effective diagnostic tool that can be readily integrated into the EHR data processing systems in primary care offices, emergency units, and hospitals alike.

4 - A Multiomics Approach to Cancer Detection

Beini Wang, Massachusetts Institute of Technology, Cambridge, MA, United States, Dimitris Bertsimas

Recent advances in high-throughput sequencing technologies have catalyzed a shift in cancer research toward multiomics. Cancer diagnosis traditionally relies on the visual inspection of histopathology slides (which are obtained from biopsies), while molecular testing is often done as a follow-up test for treatment planning and when histopathology results are inconclusive. Hence, in this work, we present a machine learning framework for robust cancer detection through a multiomics, molecular approach. Specifically, we developed parsimonious XGBoost models that integrate signals from DNA methylation profiles, miRNA expression, and gene expression (via mRNA) to distinguish tumor from normal tissue (binary classification) and identify cancer tissue of origin (multiclass classification), across diverse tumor types. Our models were trained on data from The Cancer Genome Atlas (TCGA), and validated on two datasets, namely: (i) A held-out subset from TCGA which was not used for training, and (ii) an external dataset from Phase III of the Clinical Proteomic Tumor Analysis Consortium (CPTAC-3). Multiomics models outperformed single-omics models, with the methylation-and-gene expression model yielding the highest validation AUCs. Notably, models retained strong performance using only a small number of features, enhancing explainability and potential clinical utility.

TD29

Building B Level 2 B203

Transmission Planning: Aligning Academic Advances with Industry Needs

Invited Session

ENRE: Energy

Chair: Weihang Ren, University of Florida, Gainesville, FL, United States

Co-Chair: Yongpei Guan, University of Florida, Gainesville, FL, United States

1 - A Review of Optimization Models and Challenges in Transmission Expansion Planning

Weihang Ren, University of Florida, Gainesville, FL, United States, Alinson Xavier, Yongpei Guan, Feng Qiu

Transmission expansion planning is vital in ensuring power system reliability and efficiency. Despite extensive research on optimization models for transmission planning, the lack of unified mathematical formulation and the growing complexity of academic models pose challenges for both new researchers and practical implementation. This paper reviews and synthesizes existing optimization models, proposing a unified base formulation with modular extensions to address practical considerations such as AC power flow, unit commitment, policy requirements, and uncertainty. Additionally, we survey industry technical reports to identify current planning objectives, tools, and challenges, aiming to bridge the gap between academic research and industry practice.

2 - Community-Level Optimization of Solar and Storage Investments with Energy Sharing Mechanisms

Basim Aljohani, University of Florida, Gainesville, FL, United States

In this paper, we investigate coordination and collaboration among residential property owners within a community to maximize both individual and collective profits. Recognizing that some homes may be unsuitable for solar panel installation due to factors such as tree shading, roof age, and other constraints, we propose a community-shared energy storage model. We develop three mathematical formulations (solar-only, solar with individual storage, and solar with shared community storage) and implement an ADMM-based distributed optimization framework. To handle discrete decision variables in the shared storage model, we develop a linearized reformulation with branch-and-bound algorithm. Theoretical proofs establish that solar with individual storage achieves costs no greater than solar-only systems, while shared community storage further improves upon individual storage due to economies of scale. Case studies on a 140-household Florida community demonstrate the effectiveness of the proposed approaches, with shared storage achieving cost reduction in the baseline scenario and even more improvement with Investment Tax Credits. Scalability analysis shows that cost advantages become more pronounced as community size increases and storage costs decline.

3 - Coordinated Resilience Investments for Proactive Consumers and the Supplying Utility

Tu Feng, The Ohio State University, Columbus, OH, United States, Antonio Conejo

Power grids increasingly face resilience challenges due to extreme weather and the growing penetration of weather-dependent renewable energy sources. In electricity distribution systems, both the utility and consumers seek high supply resilience. An increasing number of proactive consumers with solar panels and battery storage now enables power backups for them during utility outages. This creates a trade-off: if the utility achieves high resilience for the whole distribution system, individual consumers have little incentive to invest in their own resilience. Conversely, low utility resilience encourages more consumer investment. To address this, we propose two modeling approaches: (1) an equilibrium model, suitable when proactive consumers and the utility have comparable influence, and (2) a hierarchical bi-level model, which identifies optimal solutions for both the leading entity—the electric utility—and the responding agents, including proactive and non-proactive consumers, that react to the decisions of the leading agent. Resilience improvements are achieved through battery investment. The proposed technique can be helpful for distribution system planners to optimally coordinate resilience investments between the electric utility and proactive consumers. The complementarity models can be transformed to single-level models, although nonconvex, are still computationally tractable. We demonstrate the effectiveness of our models through case studies based on realistic distribution systems.

TD30

Building B Level 2 B204

Electricity Storage Issues: Batteries, Integration, Microgrids, and Planning

Invited Session

ENRE: Electricity

Chair: Dirk Lauinger, MIT, Cambridge, MA, United States

1 - Evaluating Energy Storage Integration Under Uncertainty: a Real Options Analysis**Gazi Nazia Nur, Iowa State University, Ames, IA, United States**

Renewable energy sources like wind power are crucial for a sustainable energy future, but they are inherently uncertain. One key challenge is the mismatch between electricity demand and wind availability: high wind generation during periods of low demand can cause grid instability and wasted wind power, while low wind during high-demand periods can lead to unmet electricity needs. The integration of energy storage systems (ESS) with wind energy farms offers a promising solution to this problem. Our study investigates the economic feasibility of integrating ESS into existing wind farms. By adding ESS, surplus energy can be stored during periods of low demand and discharged during peak hours, effectively redistributing electricity supply throughout the day and reducing peak electricity demand. This approach relieves grid congestion and lowers the locational marginal price (LMP) of electricity, ultimately reducing electricity bills for consumers. However, the high capital investment and technology costs, combined with market uncertainties, present significant challenges. To address these, we apply a real options analysis (ROA) framework, which is able to consider uncertainty and capture the value of flexibility in investment decisions. Our objective is to assess whether the benefits of electricity bill reduction justify the investment in ESS, incorporating key uncertainties such as electricity demand and storage costs. The findings aim to provide actionable insights for stakeholders considering energy storage integration in renewable energy systems under uncertain market conditions.

2 - Improved TIME-Series Clustering for Long-Term Grid Planning Using Approximate Solutions**Theodore Dahl, Massachusetts Institute of Technology, Cambridge, MA, United States, Ruaridh Macdonald**

"Energy grid planning models use time-series clustering to identify representative periods which lead to the same investment and dispatch decisions as the original, full timeseries. This reduces computational complexity and has become essential in the past few years given research showing that grid planning optimizations are only robust when they consider several decades of weather and demand data. One limitation of existing time-clustering methods is that it is difficult to know a priori how many representative periods are required to produce an accurate answer, or how long each period should be. Recently, a new clustering method was proposed which uses fast-to-produce approximate solutions as another clustering feature. This was shown to better guarantee the accuracy of the final optimization result. However, the total runtime was longer than state of the art methods and there was no way to incorporate state variables across representative periods. In this work, we build upon this recent output-based method to improve its computational efficiency while maintaining accuracy in planning outcomes. Specifically, we demonstrate that additional pre-processing steps and a second clustering step lead to fewer approximate solutions being required. We also extend the method to incorporate inter-representative period state variables, allowing long duration storage and similar technologies to be modeled. We implement the improved clustering strategies in a Julia based open-source capacity expansion model (GenX). Our results demonstrate that the optimized clustering framework can achieve promising reduction in solver runtime with only minimal deviations in investment decisions and reliability relative to the full-resolution benchmark."

3 - Infrastructure-Defining Lows In Wind and Solar Energy Resources**Shannon Hwang, Massachusetts Institute of Technology, Cambridge, MA, United States, Micah Ziegler, Jessika Trancik**

Rare, severe periods of low wind and solar resource availability could pose challenges for a transition to an electricity system that depends on substantial proportions of these variable resources. Here we examine the nature of rare but sizable resource lows to characterize the space of possible solutions for providing reliable power even during these periods. Our analysis considers the effects of variability in wind and solar energy resources over a 40-year period across the coterminous United States. We demonstrate how resource lows drive sizing of system components. For example, we show that these resource lows occur irregularly and infrequently over decadal timespans and accumulate resource deficits over timescales of days to months, heavily impacting the required capacities and utilization of energy storage, supplemental energy, and demand management infrastructure. Consequently, technologies proposed to help mitigate low resource availability, such as energy storage and supplemental (i.e., neither wind nor solar) generation, would likely operate at full capacity very infrequently, with capacity factors reaching 0.06 or less in systems with very high renewables penetrations. Resource lows can cover large geographical areas—up to hundreds of kilometers—which suggests that geographical aggregation of generation resources through transmission infrastructure expansion must be at least similarly large to be most effective. Additionally, resource lows occur during predictable seasons, which may be beneficial for the development of demand-side management strategies. Such understanding of the capacities and operations required of reliable, high-renewable energy systems is essential for informing system planning and technological innovation to address climate change.

4 - Consumer Acceptance of Second-Life use for End-of-Life EV Batteries: Evidence from a Discrete Choice Experiment in Korea**Myongji Cho, Seoul National University, Seoul, Korea, Republic of, Jungwoo Shin**

The expansion of Battery Electric Vehicles (BEVs) is regarded as a key strategy for achieving carbon neutrality and promoting energy transition. As EV adoption accelerates, the large-scale emergence of end-of-life (EOL) batteries has become an urgent environmental and societal issue. In response, two approaches have gained significant attention for their environmental and economic potential: second-life use—repurposing used batteries for applications such as energy storage systems (ESS) or low-speed electric vehicles—and resource recovery, which involves extracting valuable materials like lithium and cobalt through recycling. However, evaluating the market feasibility of these technologies requires not only technical validation but also a quantitative understanding of consumer acceptance. This study empirically examines consumer preferences and acceptance of second-life battery applications in the Korean market. A Discrete Choice Experiment (DCE) was designed to present respondents with hypothetical vehicle choice scenarios that varied in attributes such as carbon emissions, vehicle price, and battery reuse potential. A Mixed Logit Model was applied to capture heterogeneity in individual preferences. Additionally, scenario-based simulations were conducted to predict changes in demand under varying policy and information conditions. The results provide data-driven insights to support the development of demand-oriented policies and private-sector business strategies aimed at promoting sustainable management of EOL batteries in alignment with circular economy principles.

5 - From Energy Consumer to Grid Partner: Co-Designing Data Centers for System Flexibility**Wei Wang, Pacific Northwest National Laboratory, Richland, WA, United States, Himanshu Sharma, Bowen Huang, Thiagarajan Ramachandran, Shaolei Ren, Timothy Salsbury**

The rapid advancement of artificial intelligence and digital infrastructure has significantly increased the demand for data centers (DCs), which are pivotal in supporting the digital economy. Widely known for their high energy consumption, data centers now present new opportunities: with advancements in workload management strategies and appropriately sized energy storage technologies, optimally operated DCs can provide flexibility to support power grid needs. Data center developers in the early system design stage require decision-making tools that simultaneously consider design and operational decisions while accounting for workload uncertainties, ensuring demand response capabilities that meet power grid requirements. This work introduces a multi-time-scale, multi-stage stochastic control co-design framework that optimizes both design and operational decisions under uncertainty. We demonstrate our framework's application by optimizing energy storage sizing and demand response strategies for data center operations, utilizing price signals from day-ahead and real-time energy markets under uncertain workload conditions. The framework addresses operational, planning, and economic considerations, providing comprehensive solutions to co-design large loads such as data centers to enhance grid flexibility and improve resilience.

TD31

Building B Level 2 B205

Adapting Electric Distribution Systems to Climate Extremes

Invited Session

ENRE: Energy-Climate

Co-Chair: Aviadv Navon, University of Michigan, Ann Arbor, 48104, United States

Co-Chair: Michael Craig, University of Michigan, Ann Arbor, MI, United States

1 - Impacts of Rising Temperatures on Overloading and Long-Term Degradation in Electric Distribution Systems

Aviad Navon, University of Michigan, Ann Arbor, MI, United States

This work investigates how rising ambient temperatures, driven by climate change, affect the performance and longevity of electric distribution systems. Using high-resolution climate and demand data, combined with a weather-dependent network model, we quantify the increased risk of line and transformer overloading under future temperature scenarios. We also model the long-term degradation of infrastructure components due to sustained thermal stress, highlighting potential reductions in equipment lifetime and reliability. These insights support utility decision-making by helping to balance infrastructure upgrade costs with the growing risks of thermal overloading and accelerated aging.

2 - Navigating Uncertainty Through Flexibility, Resilience and New Technologies in Future Low-Carbon Power Systems**Rodrigo Moreno, University of Chile, Santiago, Chile, Mathaios Panteli**

As our energy systems face an era of unprecedented complexity driven by climate volatility and rapid decarbonization, resilience has become not just a design principle, but a necessity. This presentation explores the evolving landscape of power system resilience, highlighting key insights from global projects and groundbreaking research. Drawing on real-world applications, it will examine how integrated, risk-based, and digitally enabled strategies are redefining the way we design, operate, and plan future energy systems. From probabilistic security assessments to resilient microgrid networks, this talk makes the case for resilience not just as an add-on, but as a vital planning and operation criterion of future energy systems.

3 - Robustness and Extensibility Tradeoffs for Grid Reliability**Mikhail Chester, Arizona State University, Tempe, AZ, United States, Ryan Sparks, Nathan Johnson, Daniel Eisenberg**

Modernizing power grids while accommodating emerging technologies and confronting climate hazards necessitates new decision support frameworks that balance robustness (hardening and redundancy) and extensibility (microgrids, storage, distributed generation, etc). We introduce a novel decision support framework that: i) simulates regional grid operations during peak demand; ii) creates mechanisms for stressing the grid due to myriad hazards; iii) estimates the likelihood of cascading failures and resulting outages across transmission systems (lines and substations); iv) characterizes the marginal reliability benefits of redundancy (e.g., the addition of transmission capacity to protect critical demand nodes). The framework uses publicly available information for transmission and estimates line parameters for power flow modeling. Transmission is disrupted considering multiple hazards including wildfires and extreme heat. When disruption occurs, we rebalance power flow revealing both high likelihood low consequence and low likelihood high consequence failures at substations. We build and test this network using NERC reliability subregions. The results reveal several critical insights. First, the reliability impact to transmission systems when subjected to climate hazards and the specific vulnerabilities (as infrastructure assets and operational challenges) of these hazards. Second, the potential drivers of so-called black swan (low probability high consequence) events, as both network configuration and operational dynamics. Third, robustness and extensibility tradeoffs are captured by evaluating the marginal benefits of redundancy; portions of the grid benefit greatly from additional transmission capacity and others do not. Where the marginal benefits of additional transmission capacity are low then critical assets should be protected with extensibility strategies.

4 - Sociotechnical Modeling of Steel Electrification Interactions with Electricity Grids**Rebecca Ciez, Purdue University, West Lafayette, IN, United States, Hanwen Qin, Anindya Nath, Abhinand Ayyaswamy, Partha Mukherjee**

Steel production is one of the largest sources of industrial greenhouse gas emissions globally, and Indiana is home to most of the steel production in the United States. Many of these emissions are associated with the use of blast furnaces, which convert iron ore into steel using fuels like coal, coke, and natural gas. Electrification of the steel industry, through the use of electric arc furnaces, in conjunction with direct reduction of iron ore via hydrogen, can significantly reduce the GHG emissions per unit of steel produced. However, both electrification and the production of hydrogen by electrolysis are significant consumers of electricity.

Here, we conduct a process-based cost model to consider the tradeoffs between minimizing the cost of producing steel via electrified manufacturing, the costs to workers associated with adapting production to electricity prices, and the environmental impacts of different manufacturing strategies. To answer these questions, we use a two-stage optimization model to design and dispatch the manufacturing capacity. We use this framework to conduct scenario analysis with different grid futures and weighting of different objectives in a multicriteria decision analysis. The grid future scenarios considered are based on future technology costs, and include business-as-usual, high renewable adoption, high solar adoption, high wind adoption, and high adoption of nuclear and CCS technologies. The results of this analysis allow us to understand the appropriate size of electrified steel production facilities to reach economies of scale in steel production, and how capital, labor, and energy costs can trade-off to minimize these costs.

TD32

Building B Level 2 B206

Optimization and Impact Study of Next-Generation Energy Decarbonization Solutions

Invited Session

ENRE: Energy-Climate

Chair: Bolun Xu, Columbia University, New York, NY, United States

Co-Chair: Zhiyuan Fan, Columbia University, New York, NY, 10027, United States

1 - Consumer-Driven Climate Mitigation: An Analytical General Equilibrium Framework**Benjamin Leibowicz, The University of Texas at Austin, Austin, TX, United States**

Consumer-driven climate change mitigation strategies reduce greenhouse gas (GHG) emissions through decisions by many individual consumers to use less energy and choose less carbon-intensive products. Concrete examples include commuting by walking or cycling instead of driving, reducing air travel, living in a smaller home in a more central and walkable neighborhood, or adopting a vegetarian diet, to name just a few. The literature contains numerous studies that assess and quantify the direct GHG reduction potentials of specific consumer-driven mitigation actions, but most of this work does not consider the general equilibrium effects of these changes in final consumption and their implications for the net, economy-wide GHG impacts of the mitigation strategies. In this paper, I formulate a stylized general equilibrium model and solve it analytically to investigate the net, economy-wide GHG impacts of a shift in consumer preferences that favors less carbon-intensive final consumption. The analysis primarily serves to highlight the general channels through which changing consumer preferences influence aggregate emissions. These channels include upstream spillovers within supply chains that tend to amplify the GHG-reducing effects of climate-friendly consumption choices, as well as horizontal spillovers across sectors that tend to offset direct emissions reductions as consumers reallocate spending and prices adjust throughout the economy.

2 - Carbon-Aware Scheduling: a Framework for Integrating Intermittent Renewable Energy into Grid Operations**Walid Hanafy, UMass, Amherst, MA, United States**

The rising demand for computing, largely driven by AI applications, raises significant concerns regarding the carbon footprint of computing and the ability of the current electricity grid to accommodate this demand. Renewable energy presents a viable solution to both challenges due to its cost-effectiveness and high carbon efficiency. The widespread adoption of renewable energy sources has substantially reduced carbon intensity, measured in terms of carbon emissions per unit of energy. Nonetheless, the intermittent nature of renewable energy sources undermines their overall effectiveness as the share of renewables in an electricity grid increases. One way to enhance the effectiveness of added renewables is through demand adaptations, which align energy demand with the availability of renewable energy.

Fortunately, computing workloads have temporal and spatial flexibility that facilitates demand adaptations based on the availability of renewable energy. Carbon-aware scheduling is a framework that leverages the flexibility of computing workloads and the availability of low-carbon energy to optimize the carbon efficiency of computing. For instance, batch workloads, such as machine learning (ML) training jobs, offer substantial temporal flexibility, allowing execution to be suspended during periods of high carbon periods and scaled up again during periods of low carbon periods. Additionally, the diversity of energy sources across grids has motivated the use of spatial

shifting, which involves moving workloads according to the availability of low-cost and low-carbon energy.

In this talk, I provide an overview of the opportunities and challenges of carbon-aware scheduling, highlighting multiple use cases in batch scheduling and AI model serving systems.

3 - Optimization for Direct Air Capture System Operation and Design from Electric and Thermal Energy Perspectives

Zhiyuan Fan, Columbia University, New York, NY, United States

Direct air capture (DAC) systems are promising for large-scale carbon dioxide removal but are limited by their high electricity and thermal energy demands. This study develops a mixed-integer linear programming (MILP) framework to optimize DAC system design and operation from both power and thermal energy perspectives. We analyze two complementary strategies: (1) grid-connected DACs that dynamically respond to wholesale electricity prices, and (2) solar-thermal DACs that use concentrated solar power coupled with low-cost sand-based thermal energy storage. Grid-connected DACs can significantly improve profitability—up to a tenfold increase—by operating during low-price electricity windows, though this may reduce capacity factors. Solar-thermal DACs, in contrast, achieve high utilization (over 80%) and cost-effective CO₂ removal (\$160–\$200 per ton), particularly in regions with high solar potential and sandy terrain. Short-cycle sorbents play a critical role in both configurations by enhancing responsiveness to intermittent energy supply and ambient conditions. Together, these approaches highlight the need for integrated design strategies that leverage both grid and renewable thermal resources. Our results show that coordinated optimization of energy supply, sorbent dynamics, and policy incentives—along with careful consideration of ambient temperature and humidity sensitivity, which plays a non-trivial role in DAC performance—is essential to scale DAC systems sustainably and affordably.

4 - Underground Hydrogen Storage Opportunities Under Legacy Natural Gas Infrastructure Constraints

Wei Ai, University of Michigan, Ann Arbor, MI, United States, Jacqueline Dowling, Michael Craig

As energy systems transition to net-zero emissions, declining dispatchable fossil generation and rising variable renewable generation intensify the firming benefit of long-duration energy storage. Underground hydrogen storage (UHS), leveraging geologic formations, has extremely low energy-capacity capital costs and offers a promising solution, but its deployment is constrained geographically by resource availability. On the other hand, UHS and its supporting pipeline network can expand at lower costs and with less development efforts by leveraging existing natural gas infrastructure. To fully understand the opportunities of UHS in decarbonizing the energy system and to maximize its potential, it is essential to consider legacy natural gas infrastructure and retrofit dynamics.

We develop a coordinated energy system planning model coupling electricity, natural gas, and hydrogen sectors across the contiguous U.S. by adding a full representation of the hydrogen sector to PyPSA-USA. Modeled at the facility level, UHS and hydrogen pipeline expansions are constrained to existing natural gas infrastructures and their capacities. Energy demands for electricity, natural gas, and hydrogen are exogenously defined at the state level. The planning optimization is sequenced over multiple decision periods to capture the evolution of the energy system and UHS deployment. We also include alternative long-duration energy storage technologies to assess competition under UHS constraints.

Results will show how regional differences in subsurface resource availability and infrastructure retrofit pacing influence UHS and alternative long-duration energy storage deployment. These insights will inform strategic planning for deploying UHS under legacy natural gas infrastructure constraints.

TD33

Building B Level 2 B207

Health Policy Modeling: Advancing Decision Making for Healthcare

Invited Session

Health Applications Society

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

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

Co-Chair: Citina Liang, University of Southern California, Los Angeles, CA, United States

1 - Risk-Sensitive Optimal Acceptance for Marginal Kidneys

Yili Wang, University of Michigan, Ann Arbor, MI, United States, Mariel Lavieri, Amirhossein Moosavi

Despite nearly 90,000 patients on the US transplantation waitlist, about 25% of lower-quality (marginal) kidneys are discarded annually. To improve the utilization of these organs, an out-of-sequence allocation policy has been implemented for marginal kidneys. Motivated by this policy, this research focuses on developing decision-support models to assist patients and clinicians in making informed acceptance decisions, accounting for varying levels of risk aversion. By applying Conditional Value at Risk (CVaR) within a Markov Decision Process (MDP) framework, we capture risk sensitivity and analyze its impact on acceptance policies for marginal kidney transplants. We also demonstrate key properties of the CVaR-MDP model in determining optimal acceptance policy.

2 - Alternative Payment Models for Active Surveillance of Prostate Cancer

Sogand Soghrati Ghasbeh, University of Michigan, Ann Arbor, MI, United States, Brian Denton, Mark P. Van Oyen, Vahakn Shahinian, Arnab Srivastava

Prostate cancer is the most commonly diagnosed cancer among men in the United States aside from skin cancer. For men with low-risk prostate cancer, active surveillance (AS), a strategy that involves routine monitoring to avoid unnecessary treatment, has become the standard of care. While AS can spare patients from adverse side effects of surgery or radiation, such as incontinence and erectile dysfunction, overtreatment remains widespread. A key driver of this issue lies in the prevailing fee-for-service payment model, which incentivizes curative interventions over long-term observation. Although bundled payments have been introduced in other episodic care settings to encourage value-based treatment, such alternative payment models (APMs) do not currently exist for chronic care pathways like AS.

In this study, we develop a novel payment model to address this gap using a hidden Markov model (HMM) to characterize the natural progression of low-risk prostate cancer. Leveraging rich longitudinal data from the Michigan Urological Surgery Improvement Collaborative (MUSIC), we estimate our HMM on a contemporary cohort of patients under AS in community settings. The model captures the stochastic nature of disease evolution and treatment decisions, enabling us to simulate reimbursement distributions under current and proposed payment frameworks. We propose a payment redesign that introduces incremental reimbursements aligned with AS care timelines, aiming to make healthcare providers financially indifferent to treatment modality. This approach reduces overall healthcare spending while maintaining compensation for individual urologists, offering a promising pathway for aligning financial incentives with evidence-based care.

3 - Dynamic Assignment of Stochastic Service Requests in Mobile Public Health Operations

Jing Jin, University of Southern California, Los Angeles, CA, United States, Sze-chuan Suen, Maged Dessouky, Fernando Ordonez

For many on-site service providers, requests arrive stochastically and often must be scheduled before the following requests are observed. This is the case for on-site visits by the Mobile Vaccination Team (MVT) at the Los Angeles County (LAC) Department of Public Health. These visits provide service to a geographically disperse area with potentially long travel times. To address this problem, we formulate a finite-horizon Markov Decision Process (MDP) model where service requests arrive sequentially from random locations within the service area. Weekly nurse team schedules are built dynamically as services are requested to minimize total travel time for the next week. We study properties of the optimal policy and design heuristics for large-scale problems suitable for real-world instantiation.

4 - Strategic Influence of Household Attitudes on Hpv Vaccination in Kenya: a Network-Based Optimization Approach

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

Dissemination of health information through social networks may be an effective avenue to influence attitudes towards health interventions, as information may be received from trusted social sources. However, existing network models often neglect the possibility of heterogeneous influence and receptiveness of individuals. We

study a model to maximize the influence of Human Papillomavirus (HPV) vaccination information disseminated through a social network. Our directed network uses an extended linear threshold model to capture the dynamics of attitude spread in the presence of heterogeneous influence and receptiveness of individuals, and we provide structural observations on the equilibrium state and convergence of network outcomes. We then formulate a mixed-integer linear program (MILP) that selects a subset of nodes within the network for information dissemination to maximize the number of nodes with positive attitudes after information spreads through the social network within the analysis period. To illustrate our process, we parameterize our network model using a 2022 dataset from Kenya on household HPV vaccine attitudes. We find that the nodes selected generally consider both initial HPV attitudes and centrality measures, suggesting information on both should be included in future surveys to inform health information dissemination campaigns.

5 - Data-Driven Cohort Stratification for Markov Disease Progression Models

Nana Xia, University of Southern California, Los Angeles, CA, United States, Sze-chuan Suen

Traditional Markov models of disease progression stratify populations into cohorts by risk, demographic characteristics (age, etc.), or other factors based on data availability or ad hoc rules—practices that lack consistency, objectivity, and reproducibility. We introduce a data-driven framework that simultaneously partitions individuals into groups and estimates group-level transition probabilities in a Markov model. We formulate a mixed-integer program that minimizes the sum of squared errors between true and cohort-predicted sick-state probabilities using observed health-status trajectories. Assuming heterogeneous incidence probabilities, we solve this nonconvex problem via a coordinate-descent algorithm that alternates between optimizing individual assignments and updating group parameters. Experiments on synthetic data demonstrate that our method accurately recovers group assignments, yields interpretable transition rates, and ensures full reproducibility. This framework provides a principled, automated approach for constructing cohorts in disease-progression models, with applications in chronic disease management.

TD34

Building B Level 2 B208

Generative AI: Alignment, Reasoning and Interpretability

Invited Session

OPT: Machine Learning

Chair: Tianyi Lin, Columbia University, New York, NY, United States

1 - The Value of Imperfection: LLMs Post-Training on "Trash" Data

Peter Chen, Columbia University, New York, NY, United States

In this talk, I will discuss how to utilize "trash" data in LLM post-training. Specifically, I will focus on two areas: preference alignment and reinforcement learning (RL) for reasoning. For preference alignment, I will introduce ComPO, a framework designed to extract useful signals from "noisy" preference pairs that Direct Preference Optimization (DPO) fails to utilize effectively. ComPO employs a zeroth-order method to directly extract comparison information from the preference data and incorporate it into the model. For RL reasoning, I will present Spectral Policy Optimization (SPO), a new framework based on AI feedback that addresses the challenge of handling all-negative-sample groups in Group Relative Policy Optimization (GRPO). SPO demonstrates that reasoning mistakes can be valuable and should not be discarded in GRPO.

2 - Discrete Diffusion Models: Inference, Post-Training and Application

Hanyang Zhao, Columbia University, New York City, NY, United States

Discrete Diffusion Models is an emerging but capable family of non-autoregressive models which showcases great promise in modeling categorical distributions like language. In this talk, we propose a discrete analogue of DDIM algorithm as generalizations, which results to faster inference and improved techniques in post-training discrete diffusion models for alignment and enhanced reasoning. We also investigate the applications of discrete diffusion models for recommender systems.

3 - On Reinforcement Learning for Better Scientific Reasoning

Ming Yin, Princeton University, Plainsboro, NJ, United States

We investigate how to teach large language models (LLMs) to perform scientific reasoning by leveraging expert discussions as a learning signal. Focusing on the genomics domain, we develop an automated pipeline to extract trainable data and introduce Genome-Bench, a new benchmark constructed from over a decade of scientific forum discussions on genome engineering. Our pipeline transforms raw interactions into a reinforcement learning-friendly multiple-choice questions format, supported by 3000+ high-quality question-answer pairs spanning foundational biology, experimental troubleshooting, tool usage, and beyond. We fine-tune an LLM using RL with a rule-based reward signal derived from the synthetic MCQ dataset to enhance domain-specific reasoning.

4 - Goedel-Prover-V2

Shange Tang, Princeton University, Princeton, NJ, United States, Yong Lin, Chi Jin

We introduce Goedel-Prover-V2, an open-source language model series that sets a new state-of-the-art in automated formal proof generation. Built on the standard expert iteration and reinforcement learning pipeline, our approach incorporates three key innovations: (1) Scaffolded data synthesis: We generate synthetic proof tasks of increasing difficulty to progressively train the model, enabling it to master increasingly complex theorems; (2) Verifier-guided self-correction: The model learns to iteratively revise its own proofs by leveraging feedback from Lean's compiler, closely mimicking how humans refine their work; (3) Model averaging: We combine multiple model checkpoints to improve robustness and overall performance.

Our small model, Goedel-Prover-V2-8B, reaches 83.3% on MiniF2F test set at Pass@32, matching the performance of prior state-of-the-art DeepSeek-Prover-V2-671B while being nearly 100 times smaller in model size. Our flagship model, Goedel-Prover-V2-32B, achieves 88.1% on MiniF2F at Pass@32 on standard mode and 90.4% on self-correction mode, outperforming prior SOTA DeepSeek-Prover-V2-671B and concurrent work Kimina-Prover-72B by a large margin. Additionally, our flagship model with self-correction solves 64 problems on PutnamBench at Pass@64, securing the 1st on the leaderboard surpassing DeepSeek-Prover-V2-671B's record of solving 47 problems by Pass@1024.

TD35

Building B Level 2 B209

PSOR Best Video and Paper Competition Awards

Award Session

Public Sector OR

Chair: Hamsa Bastani, Wharton School, Philadelphia, PA, United States

Co-Chair: Maya Ganesh, Indian Institute of Management Bangalore, Bengaluru, India

Co-Chair: Rakesh Allu, Cornell University, Ithaca, NY, United States

1 - A Novel Adaptive Testing Scheme for Multi-disease Testing

Merve Bodur, University of Edinburgh, Edinburgh, United Kingdom, Douglas Bish, Ebru Bish, Berk Gorgulu

Testing for multiple diseases is essential for both diagnosis (when diseases share symptoms but require different treatments) and screening, but is highly costly. Two cost-saving strategies are multiplexing (a single assay detects multiple diseases) and pooling (a single assay tests multiple subjects, with individual re-testing if positive). While laboratory automation enables a flexible integration of these strategies, optimal testing designs that fully exploit this potential remain underexplored. Existing approaches restrict re-testing to the same multiplex used in the initial pooled test (Dorfman testing), often leading to inefficiencies. To address this gap, we introduce a novel adaptive testing design that tailors each subject's re-test based on pooled test outcomes, using only the diseases that tested positive. This leads to a complex partition-type optimization problem that is considerably more difficult to solve than the Dorfman testing designs previously studied. We derive a closed-form cost function, characterize structural properties of optimal designs, and develop effective solution methods. A case study involving 18 respiratory diseases demonstrates substantial savings: 59.8% and 53.5% relative to two current practice benchmarks, and 36.2% compared to previously studied non-adaptive Dorfman testing designs.

2 - Data-Driven Facility Layout for Enhancing Walk Appeal in Walkable Neighborhoods.

Haoran Guo, Tsinghua University, Beijing, China, People's Republic of, Sentao Miao, Wei Qi, Jian Cui, Zhenning Dong

In response to the challenges posed by car-centric urban development, there is growing global advocacy for walkable neighborhoods that promote public health and urban sustainability. To support the design of such neighborhoods, this paper studies the optimal layout of facilities (e.g., retail and service establishments) to enhance walk appeal. Unlike conventional designs that rely primarily on distance metrics, our approach focuses on pedestrian choices. We use a latent class logit model to estimate how facility layouts influence individuals' propensity to walk, accounting for heterogeneous travel purposes. This model is calibrated with a city-wide dataset from our industry collaborator, a leading Chinese map service provider. Building on this model, we develop a choice-based optimization framework to determine facility layouts that maximize the aggregate walking probability of residents. We first analyze its structural properties to derive analytical insights into optimal layout patterns. To find optimal layouts for this NP-hard problem, we reformulate it as an exponential cone program. To address the computational challenges this poses for large-scale applications, we further develop efficient greedy algorithms and establish their performance guarantees via submodularity ratio analysis. A case study in Beijing shows that our approach significantly improves walk appeal compared to traditional methods, with minimal impact on facility accessibility. Our results also reveal a "wisdom of crowds" effect in existing layouts and emphasize the importance of neighborhood context in shaping optimal design patterns. Specifically, we find that dispersed layouts are preferable in already walkable environments, while more challenging walking conditions benefit from spatially concentrated facilities.

3 - Retailers' Informational Role in an Informal Circular Economy

Amrita Kundu, McDonough School of Business, Georgetown University, Washington, DC, United States, Erica Plambeck, Qiong Wang

This paper reveals the strategic behaviour of small retailers in providing personalized prices and (mis)information about the durability of goods in an informal circular economy. Their behavior puts poor customers at an informational and economic disadvantage, and perpetuates a lemons market. These retailers sell new lead-acid batteries for electric vehicles, buy back batteries at end of life, and thus directly observe the life of batteries of various brands. Customers want long life so, in theory, a retailer will guide customers to the longest-lasting brand for which the marginal cost is well justified, to build reputation and attract a stream of repeat customers. To observe how retailers actually behave, we covertly audit 261 retail shop owners. Posing as five different types of customer across 1,122 visits, we learn that retailers provide inconsistent information about battery life and which brand is longest-lasting. For poor customers, a retailer exaggerates the life and sets a higher price (net of used battery trade-in value) for a given brand of battery. More frequently with poor customers, a retailer promotes as longest-lasting the brand that offers him the best incentive. A retailer gives the lowest price for a given brand of battery to a wealthy customer who owns multiple vehicles. The retailers do not create a secondary market for well-performing used batteries. Instead, retailers offer to buy used battery for about 1/2 the price of new battery, regardless of its age or performance, and most retailers simply sell the used batteries to informal battery scrapers.

4 - Improving Basic Retail Coverage in Africa: Modeling Geographic Disadvantage Among Farmers in Senegal

Mark Brennan, Rutgers University, Camden, NJ, United States, Stephen Graves, Peter Zhang, Babacar Ndiaye, Jonars Spielberg, Bish Sanyal

African consumers rely on 10 million retail points for essentials: medicine, schoolbooks, and farming equipment, among others. The lack of accessibility and competitiveness of the retail landscape for significant portions of the population is a salient policy concern. Due to a historical lack of data, the questions in African capitals today are basic: where does and should retail exist relative to disadvantaged customer segments? We use an original census of 1,100 irrigation-equipment shops along the 600-kilometer Senegal River Valley, combined with satellite imagery that identifies where demand for irrigation products lies. We find that farmers need to travel an average distance of 31 kilometers to buy any irrigation equipment, and half of these farmers lack a shop in their nearest city. The retail sector is particularly sparse in areas where the state also delivers fewer public services. However, we find that if policymakers were to induce just five more shops to open in prescribed locations, the farmers' mean travel distances to buy most equipment types would be lessened by one-third—and especially improved for farmers underserved by the state. Our results provide policymakers with direction to locate new retail irrigation equipment shops and to encourage more prevalent repair shops to also sell equipment. Our approach of combining field and satellite data can be applied in other contexts to inform retail-policy decisions. Our hope is that this work generates awareness about geographic disadvantage that African farmers face and demonstrates that modest policy action can correct much of this disadvantage.

TD36

Building B Level 2 B210

Recent Advances in Stochastic Nonlinear Optimization I

Invited Session

OPT: Nonlinear Optimization

Chair: Miaolan Xie, University of Chicago, Chicago, IL, United States

Co-Chair: Sen Na, Georgia Institute of Technology, Atlanta, 30339, United States

1 - Advanced, Adaptive and Flexible Algorithms for Decentralized Optimization

Albert Berahas, University of Michigan, Ann Arbor, MI, United States, Raghu Bollapragada

Decentralized optimization problems come in diverse shapes and forms, and could have very different characteristics. In this talk, we discuss novel flexible approaches for solving decentralized optimization problems that adapt to problem characteristics. We discuss the rationale behind our proposed techniques, convergence in expectation and complexity guarantees for our algorithms, and present encouraging numerical results. This is joint work with Raghu Bollapragada, Wanping Dong, Shagun Gupta, Suhail Shah and Jiahao Shi.

2 - Reducing Sample Complexity in Stochastic Derivative-Free Optimization via Tail Bounds and Hypothesis Testing

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

We introduce and analyze new probabilistic strategies for enforcing sufficient decrease conditions in stochastic derivative-free optimization, with the goal of significantly reducing sample complexity and simplifying convergence analysis. In particular, we propose two complementary approaches. First, we develop a new tail bound condition imposed on the estimated reduction in function value, which permits flexible selection of the power used in the sufficient decrease test—greater than 1 but arbitrarily close to it. This approach allows us to reduce the number of samples per iteration from the standard $O(\delta^{\alpha-4})$ to $O(\delta^{\alpha-2-\epsilon})$, where $\epsilon > 0$ can be made arbitrarily

small, under mild noise assumptions. Second, we formulate the sufficient decrease condition as a sequential hypothesis testing problem, in which the algorithm adaptively collects samples until the evidence suffices to accept or reject a candidate step. This test provides statistical guarantees on decision errors and can further reduce the required sample size, particularly in the Gaussian noise setting, where it approaches $O(\delta^{-2})$ when the decrease is far from the "forcing function". We incorporate both techniques into stochastic direct-search and trust-region methods for potentially non-smooth, noisy objective functions, and establish their global convergence properties.

3 - Fast, Adaptive, and Scalable Methods via Hessian Averaging and Adaptive Gradient Sampling Strategies

Raghu Bollapragada, The University of Texas at Austin, Austin, TX, United States, Thomas O'leary Roseberry

In this talk, we discuss minimizing finite-sum and expectation objective functions using Hessian-averaging-based subsampled Newton methods. These methods accommodate gradient inexactness and maintain fixed per-iteration Hessian approximation costs. Recent work (Na et al. 2023) showed that Hessian averaging achieves fast $\mathcal{O}(\sqrt{\frac{\log k}{k}})$ local superlinear convergence for strongly convex functions, but requires gradient exactness and strong convexity, limiting practical use. To address this, we propose Hessian-averaged methods with adaptive-sampling strategies allowing gradient inexactness. For finite-sum problems, we use deterministic sampling, yielding global linear and sublinear convergence for strongly convex and nonconvex functions. We derive an improved local superlinear rate of $\mathcal{O}(\frac{1}{\sqrt{k}})$. For expectation problems, we use stochastic sampling and derive global linear/sublinear rates and a local superlinear rate of $\mathcal{O}(\frac{1}{\sqrt{k}})$. Additionally, we introduce scalable methods like the diagonally-averaged Newton (Dan) method for large-scale problems. Numerical results show that Hessian averaging enhances convergence and achieves state-of-the-art performance on challenging tasks like CIFAR100 classification with ResNets.

4 - Variance-reduced First-order Methods for Stochastic Optimization with Deterministic Constraints

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

We consider stochastic optimization problems with deterministic constraints. Existing methods typically focus on finding an approximate stochastic solution that ensures the expected constraint violations and optimality conditions meet a prescribed accuracy. However, such an approximate solution can possibly lead to significant constraint violations in practice. To address this issue, we propose variance-reduced first-order methods that treat the objective and constraints differently. Under suitable assumptions, our proposed methods achieve stronger approximate stochastic solutions with complexity guarantees, offering more reliable constraint satisfaction than existing approaches.

TD37

Building B Level 2 B211

Ongoing Advances in Decision-Aware Learning

Invited Session

OPT: Optimization Under Uncertainty

Chair: Michael Huang, CUNY Baruch College, New York, NY, United States

1 - Statistical Comparative Analysis of Data-Driven Optimization Algorithms

Haofeng Zhang, Morgan Stanley, New York, NY, United States, Adam Elmachtoub, Henry Lam, Haixiang Lan

Data-driven optimization aims to translate a machine learning model into decision-making by optimizing decisions on estimated costs. Such a pipeline can be conducted by fitting a distributional model which is then plugged into the target optimization problem. While this fitting can utilize traditional methods such as maximum likelihood, a more recent approach uses estimation-optimization integration that minimizes decision error instead of estimation error. Although intuitive, the statistical benefit of the latter approach is not well understood yet is important to guide the prescriptive usage of machine learning. In this paper, we dissect the performance comparisons between these approaches in terms of the amount of model misspecification. In particular, we show how the integrated approach offers a "universal double benefit" on the top two dominating terms of regret when the underlying model is misspecified, while the traditional approach can be advantageous when the model is nearly well-specified. Our comparison is powered by finite-sample tail regret bounds that are derived via new higher-order expansions of regrets and the leveraging of a recent Berry-Esseen theorem.

2 - Decision-Focused Bias Correction for Fluid Approximation

Can Er, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States, Mo Liu

Staffing decisions in large-scale service systems, such as call centers and healthcare networks, are typically formulated as a two-stage stochastic optimization problem: the first stage determines the capacity of each server, and the second stage dynamically allocates different types of customers to each server. These decisions are generally made in two steps: estimating future demand for different customer types and optimizing capacity accordingly. Due to the unknown joint distribution of demand, a common approach is to use fluid models that consider average arrival rates to solve the two-stage problem. Despite the widespread use of the fluid approximation (FA), employing the mean (or conditional expectation) as the arrival rate has been the conventional practice.

In our paper, we show that the traditional FA approach, which uses the sample mean of the arrival rate, does not generally yield the optimal staffing decision. This motivates a corrected approach that directly minimizes the expected cost. We characterize the resulting predictor and show that it can be computed as the minimizer of a convex function. We then analyze how this predictor can be estimated from data and establish convergence guarantees for gradient-based optimization, along with generalization error bounds under finite samples. Finally, numerical experiments demonstrate the effectiveness of our bias correction approach for FA under various settings.

3 - Decision-Aware Training of Spatiotemporal Forecasting Models to Select a Top K Subset of Sites for Intervention

Kyle Heuton, Tufts University, Medford, MA, United States, F. Samuel Muench, Shikhar Shrestha, Thomas J. Stopka, Michael C. Hughes

Optimal allocation of scarce resources is a common problem for decision makers faced with choosing a limited number of locations for intervention. Spatiotemporal prediction models could make such decisions data-driven. A recent performance metric called fraction of best possible reach (BPR) measures the impact of using a model's recommended size K subset of sites compared to the best possible top-K in hindsight. We tackle two open problems related to BPR. First, we explore how to rank all sites numerically given a probabilistic model that predicts event counts jointly across sites. Ranking via the per-site mean is suboptimal for BPR. Instead, we offer a better ranking for BPR backed by decision theory. Second, we explore how to train a probabilistic model's parameters to maximize BPR. Discrete selection of K sites implies all-zero parameter gradients which prevent standard gradient training. We overcome this barrier via advances in perturbed optimizers. We further suggest a training objective that combines likelihood with a decision-aware BPR constraint to deliver high-quality top-K rankings as well as good forecasts for all sites. We demonstrate our approach on two where-to-intervene applications: mitigating opioid-related fatal overdoses for public health and monitoring endangered wildlife.

4 - Decision-Aware Learning with Feature-Space Clustering Constraints

Golbagh Roghaniaraghi, Baruch College, NEW YORK, NY, United States, Michael Huang

In many real-world decision-making tasks, such as bank marketing, fraud detection, and customer churn, predictive models are trained on high-dimensional data that often contain weakly informative or noisy features. These features may not harm prediction accuracy but can mislead downstream decision-making when predictions are used in combinatorial optimization. We propose a decision-aware learning framework that addresses this challenge by using unsupervised clustering to denoise the feature space in a way that improves decision quality. Specifically, we introduce clustering constraints into the optimization layer, requiring that similar instances, grouped via hierarchical clustering, are treated consistently in downstream decisions. Unlike traditional clustering approaches, we do not fix the clusters in advance. Instead, we tune the clustering structure based on its impact on decision loss, as measured by the Perturbation Gradient (PG) loss. Empirical results across all three domains show that our method leads to improved decision quality compared to standard predict-then-optimize models, as well as PG-trained models without clustering constraints. By integrating

unsupervised clustering into a decision-aware optimization pipeline, our approach enables more reliable policies in settings where noisy features would otherwise obscure important structure in the data.

5 - Bandit-Powered Adaptive Decision Making

Huiwen Jia, University of California, Berkeley, Berkeley, CA, United States, Haoming Shen

We study adaptive decision-making problems under uncertainty, where decisions unfold in stages and the decision-maker selects a primary discrete decision followed by an auxiliary linear or mixed-integer decision in each stage. A key aspect of this setting is that the distribution of the uncertain parameters depends solely on the primary decision. To address this structure, we propose a bandit-based framework that guides exploration over the primary decision and leverages observed feedback to optimize the auxiliary decision. This approach enables effective learning in layered decision settings with decision-dependent uncertainty. We establish theoretical guarantees on the performance of the proposed method.

TD38

Building B Level 2 B212

Derivative Free and Simulation Optimization

Invited Session

OPT: Computational Optimization and Software

Chair: Juliane Mueller, National Renewable Energy Laboratory, Golden, CO, United States

Co-Chair: Yunsoo Ha, National Renewable Energy Laboratory, Raleigh, United States

1 - Remu: Regional Minimal Updating for Model-Based Derivative-Free Optimization

Pengcheng Xie, Lawrence Berkeley National Laboratory, Berkeley, CA, United States, Stefan Wild

Derivative-free optimization (DFO) problems are optimization problems where derivative information is unavailable or extremely difficult to obtain. Model-based DFO solvers have been applied extensively in scientific computing. Powell's NEWUOA (2004) and Wild's POUNDERs (2014) explore the numerical power of the minimal norm Hessian (MNH) model for DFO and contributed to the open discussion on building better models with fewer data to achieve faster numerical convergence. Another decade later, we propose the regional minimal updating (ReMU) models, and extend the previous models into a broader class. This talk shows motivation behind ReMU models, computational details, theoretical and numerical results on particular extreme points and the barycenter of ReMU's weight coefficient region, and the associated KKT matrix error and distance. Novel metrics, such as the truncated Newton step error, are proposed to numerically understand the new models' properties. A new algorithmic strategy, based on iteratively adjusting the ReMU model type, is also proposed, and shows numerical advantages by combining and switching between the barycentric model and the classic least Frobenius norm model in an online fashion.

2 - Adaptive Sampling-Based Bi-Fidelity Stochastic Trust Region Method for Derivative-Free Stochastic Optimization

Yunsoo Ha, National Renewable Energy Laboratory, Golden, CO, United States, Juliane Mueller

Bi-fidelity stochastic optimization is increasingly favored for streamlining optimization processes by employing a cost-effective low-fidelity (LF) function, with the goal of optimizing a more expensive high-fidelity (HF) function. We introduce ASTRO-BFDF, a new adaptive sampling trust region method specifically designed for solving unconstrained bi-fidelity stochastic derivative-free optimization problems. Within ASTRO-BFDF, the LF function serves two purposes: first, to identify better iterates for the HF function when a high correlation between them is indicated by the optimization process, and second, to reduce the variance of the HF function estimates by Bi-fidelity Monte Carlo (BFMC). In particular, the sample sizes are dynamically determined with the option of employing either crude Monte Carlo or BFMC, while balancing optimization error and sampling error. We demonstrate that the iterates generated by ASTRO-BFDF converge to the first-order stationary point almost surely. Additionally, we numerically demonstrate the superiority of our proposed algorithm by testing it on synthetic problems and simulation optimization problems with discrete event simulations.

3 - A Large-Scale Benchmarking of Deterministic and Stochastic Black-Box Optimization Algorithms

Donghyun Oh, Georgia Institute of Technology, Atlanta, GA, United States, Nikolaos Sahinidis

We present a systematic benchmarking of state-of-the-art black-box optimization algorithms. Using a diverse and extensive set of test problems, ranging from smooth and convex to nonsmooth and nonconvex, we evaluate optimization performance with metrics including solution quality, convergence rate, and sensitivity to initialization. In addition to customary stochastic methods such as Bayesian optimization, we consider a variety of modern deterministic approaches like the recently proposed branch-and-model algorithm. We report trends in performance and scalability, highlighting the strengths and limitations of different algorithms across multiple problem classes.

4 - Global Multi-Objective Simulation Optimization: An Upper Bound on the Error with Convergence Analysis for a Simple Algorithm

Susan Hunter, Purdue University, West Lafayette, IN, United States, Burla Ondes

Consider the context of solving a multi-objective simulation optimization problem with two or more simultaneous and conflicting continuous objective functions to global optimality on a compact feasible set. If computing resources are plentiful or the simulation model runs quickly, one might implement a simple algorithm which consists of selecting m feasible points in the decision space using a space-filling design, expending n simulation replications per point to estimate the objective function values, and returning the discretized estimated efficient and Pareto sets. In this context, we define a mathematically tractable optimality gap for assessing the algorithm convergence as the total simulation budget $b=mn$ tends to infinity, derive probabilistic upper bounds on the optimality gap of the returned solution, and determine how best to trade off the number of feasible points m with the number of simulation replications per point n to ensure convergence of the optimality gap to zero at a fast rate.

5 - The Derivative-Free Fully-Corrective Frank-Wolfe Algorithm for Optimizing Functionals over Probability Spaces

Di Yu, Purdue University, West Lafayette, IN, United States, Shane Henderson, Raghu Pasupathy

The challenge of optimizing a smooth convex functional over probability spaces has recently generated much attention due to its relevance in important contexts such as experimental design, emergency response, and variations of the problem of moments. For solution, the fully-corrective Frank-Wolfe (FCFW) first-order recursion has emerged as a viable and provably efficient algorithm. We propose an FCFW recursion that rigorously handles the oft-encountered zero-order setting, where the derivative of the objective is known to exist, but only the objective at a specified probability measure is observable. The main ingredient of our proposal is an estimator for the objective's influence function, which can be viewed as a function providing the directional derivative of the objective function in the direction of point mass probability distributions, constructed via a combination of Monte Carlo, and a projection onto the orthonormal expansion of an L2 function on a compact set. An analysis of the bias and variance of the influence function estimator guides step size and Monte Carlo sample size choice, and leads to a characterization of the recursive rate behavior on smooth non-convex problems.

TD39

Building B Level 2 B213

Exploiting Math Structure for Faster Solve Times on Practical Problems

Invited Session

OPT: Integer and Discrete Optimization

Chair: Alexandra Newman, Colorado School of Mines, Golden, CO, United States

1 - Improvements and Enhancements to Modeling Optimized Steel Coil Production Schedules Under Continuous Casting and Hot Rolling

Jennifer McClary, Colorado School of Mines, Colorado Springs, CO, United States, John Cox, Alexandra Newman

In continuous steel casting, heats of molten steel are alloyed, cast into slabs, and hot-rolled into coils in a tightly coordinated process. Building on a previous mixed-integer programming model for daily scheduling in a 100% direct-charge steel mill with two casters and one hot rolling mill, we present an enhanced model that improves

computational scalability while preserving solution quality. The extended model explicitly defines penalty parameters for violating operational best practices and introduces three constraint categories: intercaster gauge differences, gauge change, and double trim operations. Three innovations further enhance efficiency: (1) learned cuts derived from production insights to tighten the formulation, (2) coil clustering to reduce sequencing complexity, and (3) iterative solving that prioritizes critical penalties for faster convergence via warm starts. Computational experiments demonstrate that the model solves large instances with thousands of variables significantly faster than the original, enabling real-time scheduling and future planning in steel mills. These advances leverage state-of-the-art solvers and domain-specific insights, addressing the combinatorial complexity of large-scale schedules while maintaining strict safety and logical constraints

2 - Optimizing Freight Logistics for Mining Operations Under Infrastructure Constraints

Yongchen Wu, Colorado School of Mines, Golden, CO, United States, Kevin Bamwiso, Steph Shiferaw, Hojun Choi, Alexandra Newman

We analyze an integrated freight logistics system for mining operations in the United States. Our study focuses on optimizing the fleet size and cost over multiple years at monthly fidelity, within a constrained multi-modal (truck and rail) transportation network.

We formulate and solve a multi-commodity, time-expanded network flow model that minimizes transportation operational costs while satisfying material needs for mining and processing operations. The model captures dynamic supply and demand, transport capacity constraints, storage limits, and commercial market purchases.

We provide quantitative trade-offs between transport modes and offer recommendations on acid transloading, storage policies, and routing protocols for the time horizon of interest.

3 - Tightening Bilinear Relaxations with Geometric Rotations

Jaivardhan Sood, The University of Texas at Arlington, Arlington, TX, United States, Jay Rosenberger

Bilinear terms are crucial in nonlinear optimization, presenting considerable challenges due to their non-convexity. While traditional techniques use McCormick envelopes to form convex relaxations, they often result in loose bounds, particularly when variable limits create an elongated rectangular shape. This research introduces a new geometric approach to improve these relaxations. By analyzing the feasible region's geometry, we propose a 45-degree rotation of the coordinate system when one edge of the bounding rectangle is significantly larger than the other. This rotation facilitates the derivation of directional bounds that consistently surpass the upper bounds given by conventional McCormick envelopes. Instead of relying on the typical axis-parallel cuts from McCormick relaxations, our method develops 45-degree cuts that more efficiently tighten the relaxation. We show that this approach yields better upper bounds in all scenarios, representing a notable improvement for global optimization solvers, especially within spatial branch-and-bound frameworks.

4 - A Branch-and-Price Algorithm to Solve the Two Dimensional Cutting Stock Problem with Length Tolerances

Serkan Urkmez, Ozyegin University, Istanbul, Turkey, Erhun Kundakcioglu

The Two-Dimensional Cutting Stock Problem (TDCSP) poses a critical optimization challenge in industries such as flat steel, textile, glass, paper, and wood processing, where maximizing material utilization is essential for minimizing waste and reducing costs. This study introduces a mixed-integer linear programming model for TDCSP with guillotine cuts, incorporating industry-specific constraints such as order specifications defined by length intervals, precise width requirements, and total demanded lengths. Given the exponential growth in computational complexity, we employ column generation as the primary solution approach, systematically generating and integrating cutting patterns to improve efficiency. To further enhance performance, we incorporate a branch-and-price algorithm, applying heuristics and warm-start strategies to accelerate convergence. Computational experiments on large-scale instances demonstrate that column generation with branch-and-price outperforms the direct solution of the standard formulation, significantly reducing material waste while ensuring practical applicability. These results highlight the superior scalability, adaptability, and effectiveness of the proposed approach, delivering both economic and environmental benefits for the industry.

5 - Mixed-Integer Optimization for Causal Learning with Unknown Interventions

Tong Xu, Northwestern University, Evanston, IL, United States, Simge Kucukyavuz, Armeen Taeb, Ali Shojaie

Causal discovery is an important problem in many areas such as biology, social sciences, and economics. There have been a lot of methods using observational and interventional data to learn the Bayesian network that represents the interaction network among a complex system. Current methods for causal discovery often face one or more of the following limitations: (i) reliance on known intervention targets, restricting their applicability to specific scenarios (e.g., GIES); (ii) lack of scalability and dependence on restrictive parametric assumptions, often yielding only partial causal structures; (iii) reliance on test-based approaches that require large sample sizes, making them impractical for data-scarce settings; and (iv) inconsistency in their results. These shortcomings underscore the need for more scalable, flexible, and consistent methods for causal discovery.

We propose a novel mixed-integer programming formulation for causal structure learning that unifies observational and interventional settings, accommodating known, partially known, and unknown intervention targets while incorporating structural information. We develop an efficient alternating optimization algorithm with guaranteed convergence to a coordinate-wise minimum. Empirical results demonstrate that our method achieves superior scalability and accuracy compared to state-of-the-art algorithms, bridging a critical gap in causal discovery and providing a robust foundation for future research on scalable DAG learning.

TD40

Building B Level 2 B214

Service Science II

Contributed Session

Contributed

Chair: Shrutivandana Sharma, New York University, Brooklyn, NY, United States

1 - Work-from-Home and Its Implications for Firm Performance

Yu Wang, Binghamton University, Vestal, NY, United States, Rory Eckardt

With the advancement of technologies and the outbreak of COVID-19, work-from-home (WFH) has transitioned from an alternative to a dominant work mode. While extensive extant literature has emphasized individual-level outcomes of it, this study investigates its actual effects on the organizational level, focusing on how different levels of WFH intensity (in-office, hybrid, fully remote) influence productivity and profitability in professional service firms.

Drawing on resource based view and social exchange theories, we propose that WFH can enhance firm performance by reducing operational costs, increasing autonomy, and expanding access to a broader talent pool and a wider clients base. In addition, we further examine three moderators that shape this relationship: the gender of strategic leaders, the firm's leverage ratio (i.e., number of associates per partner), and the level of AMO human resource systems. Female leaders may amplify WFH benefits through relational leadership styles; high leverage ratios—associated with structured workflows—may support more efficient remote collaboration; and strong AMO systems enhance employee performance under remote settings.

Based on archival data from ALM 200 law firms, we empirically test these relationships by using OLS and moderated hierarchical regression analyses. This research offers new insights into when and for whom WFH is most effective, advancing theoretical understanding of remote work and providing actionable guidance for strategic decision-making in professional service firms.

2 - Contract Design for Professional Services

Jiayu Tian, Washington State University, Pullman, WA, United States, ting lu

Professional service contracts often face expertise asymmetry. We develop a performance-based contract framework, which improves alignment between client and provider incentives and enhances contract efficiency. This study advances our understanding of contract theory and practice.

3 - Queue-Length Information in a Tiered Service System with Referrals and Imperfect Customer Perceptions

Shrutivandana Sharma, New York University, Brooklyn, NY, United States, Yufeng Zhang

In this work we investigate implications of queue-length information provision on strategic customers' choice between tiered service providers, where customers are heterogeneous in terms of their service needs and the generalist tier refers urgent customers to the expert tier, but where customers' choices are made under imperfect perceptions of their urgency (need for referral). We present an analytical framework based on queueing games to investigate these issues. We prove existence of equilibrium for this system, and characterize structure of equilibrium strategies. Based on this structure we conduct numerical experiments to further analyze equilibrium properties, and implications of queue-length information on social cost for customers. We also compare this system with one without real time queue-length information, and identify conditions under which queue-length information provision is beneficial.

TD41

Building B Level 3 B301

Advancing Healthcare Using Large Language Models (LLMs)

Invited Session

Health Applications Society

Chair: Nahian Tahmin, UTK, K, TN, United States

Co-Chair: Anahita Khojandi, University of Tennessee, Knoxville, TN, United States

Co-Chair: Nahian Tahmin, The University of Tennessee Knoxville, Knoxville, TN, United States

1 - Assessing the Applicability of Large Language Models in the Multi-Attribute Healthcare Decision Making

Mohammad Alipour Vaezi, Virginia Tech, Blacksburg, VA, United States, Sajad Khodadadian, Kwok-Leung Tsui

The growing complexity of healthcare decision-making has underscored the necessity for robust, scalable, and interpretable decision-support systems. Multi-Attribute Decision-Making (MADM) techniques have been widely employed to facilitate structured decision processes, enabling the evaluation of alternatives across multiple, often conflicting criteria. Traditionally, these methods rely heavily on expert judgment to define decision criteria, assess attribute importance, and rank alternatives. Recent advancements in Large Language Models (LLMs) offer a promising avenue to augment or enhance MADM applications in healthcare by leveraging their natural language processing capabilities to synthesize information, generate structured decision matrices, and provide analytical insights. This study proposes a novel framework that integrates LLMs into MADM processes, aiming to assess their efficacy in replicating expert-driven decision-making tasks. The framework includes prompt-tuning strategies and meta-learning techniques to adapt LLM outputs to MADM requirements, enabling automated identification of decision criteria, construction of decision matrices, and application of established MADM methods. To evaluate this approach, three case studies are designed to benchmark LLM-driven decision-making against expert-based evaluations, focusing on key performance metrics such as consistency, interpretability, biasness, and alignment with ethical guidelines. Preliminary findings suggest that LLMs possess the capacity to enhance decision consistency and scalability while maintaining interpretability within MADM frameworks. This study contributes to the discourse on AI-driven decision-support in healthcare, offering insights into the practical integration of LLMs in expert-based decision-making environments.

2 - Toward the Effectiveness and Efficiency of LLM Reasoning in Healthcare

Liang Zhao, Emory University, Atlanta, GA, United States

As medical AI systems increasingly support high-stakes decisions, it is no longer sufficient for models to be accurate—their decision process must also be transparent, reasonable, adaptable, and resource-efficient. In this talk, I will introduce a roadmap on enhancing AI reasoning, covering techniques like explanation supervision, distillation, and retrieval-augmented generation (RAG). First, explanation supervision allows models to learn not just *what* to predict, but *why*, using human or saliency-based guidance to align model reasoning with clinical intuition. Building on this, we will introduce the reasoning of large language models (LLMs) and how to improve its quality and efficiency on the fly. Then, we extend to the topic on how large language models (LLMs) can act as mentors in an *explanation-guided distillation* process—actively correcting the reasoning flaws of smaller, deployable models, while minimizing annotation costs. Finally, we address the need for comprehensive context in clinical decision support by extending RAG methods to go beyond isolated documents and tap into richly connected patient data, biomedical literature, and knowledge graphs. Together, these approaches move us toward a future where medical AI is not only performant, but also interpretable, scalable, and grounded in structured clinical knowledge. This integrative strategy opens new doors for trustworthy AI applications in diagnosis, imaging, and real-time decision-making in the clinic.

3 - Paging Dr. GPT: Extracting Information from Clinical Notes to Enhance Patient Predictions

David Anderson, Villanova School of Business, Villanova, PA, United States, Margret Bjarnadottir, Stephen Mahar, Shriyan Reyya

In healthcare, there is a long tradition of using information from electronic medical records (EMRs) for outcome prediction in support of hospital operation and to support clinical decision making. Most prior work uses tabular data from EMRs and overlooks the unstructured clinical notes typically written by doctors. These notes capture key aspects of the patient's hospital stay, including diagnosis, treatment, progress, medications and care plans, which are factors important to the patient's health and prognosis that may contribute to prediction accuracy. Despite their potential, these notes remain underutilized in predictive modeling, even though large language models (LLMs), with their strengths in text summarization and analysis would seem well suited for extracting insights from them. Here we introduce an intuitive, physician-driven approach that integrates LLMs into standard machine learning pipelines, and show that a model using an LLM's answers to simple clinical questions from discharge notes can outperform models built on available tabular data. We further show that combining the LLM's answers with tabular data significantly improves predictive performance. Our results reveal that LLMs are capable of extracting and distilling the information contained in discharge notes to quantify patient risks, and illustrate a way that AI tools can complement human workers. This approach highlights the potential for transparent and expert driven prompt engineering to combine the strengths of human expertise and machine intelligence, and is broadly applicable in any field with underutilized textual data sources, where LLMs can be guided with expert questioning.

TD42

Building B Level 3 B302

Digital Health Innovation

Invited Session

Health Applications Society

Chair: Yue Hu, Stanford University, Stanford, CA, United States

1 - Regression Discontinuity in Time: Evaluating the Impact of Evolving Digital Health Interventions

Isha Thapa, Stanford University, Stanford, CA, United States, Pierre-Amaury Laforcade, Franziska Bishop, Johannes Ferstad, Manisha Desai, David Maahs, Priya Prahalad, Dessi Zaharieva, David Scheinker, Ramesh Johari

Background: Clinics continue to adopt digital health interventions (DHIs) in which algorithms analyze data to help direct patient care. Changes to these algorithms are rarely evaluated rigorously, despite their potential to affect patient outcomes. The regression discontinuity in time (RDT) design may be used to estimate the causal effect

of such changes but has received little attention in medical literature.

Methods: We conducted a retrospective study of continuous glucose monitor (CGM) data from youth with type 1 diabetes in the 4T Program from November 2020 to March 2022. In 2020, the clinic used an algorithm that directed patients for provider review based on the number of glucose targets not met (e.g., time in range (TIR) < 65%). In September 2021, the clinic adopted a new algorithm that prioritized the review of patients experiencing level 2 hypoglycemia (glucose < 54 mg/dl). We evaluated the validity of the RDT framework in this setting and estimated the impact of this change in directed care on weekly TIR.

Results: There were 247 patients with 11,297 weekly TIR observations. Robustness checks supported the validity of the RDT design. Patients with level 2 hypoglycemia were prioritized for review, with no significant change to population-level TIR (-0.2% points; 95% CI: [-4.8, 3.5]).

Conclusions: We demonstrate the feasibility of using the RDT framework to estimate the impact of algorithmic changes to a DHI on patient care. As algorithm-directed care increases in clinical practice, this approach can serve as a diagnostic tool to measure how operational changes impact outcomes.

2 - Evaluating Secure Messaging Workflows to Improve Healthcare Delivery

Sean Tsung, Stanford University, Stanford, CA, United States, Jisun Lee, Yue Hu, Song-Hee Kim, Vijaya Parameswaran

Secure messaging is widely used for asynchronous patient-provider communication, yet the effects of messaging workflows on both operational and patient outcomes remain understudied. We analyze secure messaging data from Stanford primary care clinics to assess how message routing strategies and varying levels of resource flexibility (e.g., increased involvement of medical assistants) influence key operational outcomes such as response times and resolution rates, along with downstream patient outcomes. While greater provider flexibility may be associated with shorter response times, it can also be linked to lower resolution rates. Our findings offer guidance for designing secure messaging workflows and managing provider workloads to improve system performance without compromising care quality.

3 - Personalized Health Check with Joint Screening: a Data-Driven Approach with Pomdp

Guohua Wan, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Haotian Li

Regular health check is critical in personal health management, with over 250 million conducted annually in China. Despite its widespread use, designing an efficient health check policy remains a challenge. Existing clinical guidelines often fail to offer personalized recommendations and overlook the influence of disease correlations on joint disease screening.

We address this issue by developing a stochastic modeling framework based on Partially Observable Markov Decision Processes to personalize health check decisions. The model decides optimally when and how to select a subset of screening items for an individual, explicitly accounting for correlations between diseases. To estimate the parameters of the decision model, particularly the joint progression of multiple diseases, we introduce a modified Baum-Welch algorithm and apply it to real-world health check data. The results are validated against medical literature in our numerical studies. In the case study, we demonstrate the application of our personalized health check strategy by designing a joint screening strategy for chronic kidney disease (CKD) and diabetes. The proposed policy is projected to save 628 million CNY annually in China compared to current guidelines and 558 million CNY compared to single-disease analysis. Additionally, it could save up to 6.828 quality-adjusted life months for the stratified population compared to existing screening benchmarks. Our findings highlight that the proposed personalized health check policy offers more efficient and cost-effective recommendations, with significant potential for improving health outcomes and resource allocation.

4 - Telehealth's Double-Edged Sword: The Hidden Costs of Expanding Access to Virtual Care

Masoud Kamalahmadi, University of Miami, Coral Gables, FL, United States, Jong Myeong Lim

Leveraging variation in telehealth adoption across specialties during the COVID-19 pandemic within a major health system, we find that the expanded access offered by telehealth may come with significant trade-offs, including increased physician workload, reduced patient time, and financial strain on healthcare organizations.

TD43

Building B Level 3 B303

Emerging Topics at the Interface of Finance, Operations, and Risk Management

Invited Session

MSOM: iForm

Chair: Jiri Chod, Boston College, Chestnut Hill, MA, United States

Co-Chair: Jiri Chod, Boston College, Chestnut Hill, MA, United States

1 - Auditing Strategies in Trade Credit Contracts

Paola Martin, Indiana University, Bloomington, IN, United States

In a trade credit contract, the supplier may not observe the retailer's marketing efforts and resulting sales. Verification of the retailer's sales is possible, but costly. Furthermore, the supplier may not accurately learn the true sales. The supplier may choose to perform an early audit, which is less costly and inaccurate, and/or a late audit, which is more costly but reveals the actual sales of the retailer. In either audit, the supplier does not learn the retailer's marketing effort. This talk explores the supplier's choice of auditing strategies and their impact on trade credit contract terms. In particular, it will address the following questions. Will the supplier exercise the option to perform an early audit? Given the results of the early audit, for what parameter values will the supplier forego a second audit? How does the auditing strategy affect the retailer's order quantity, as well as the total supply chain profit? How does the asset level of the retailer affect the supplier's trade credit terms?

2 - Clean-Tech Adoption and Risk Hedging

Liao Wang, The University of Hong Kong, Hong Kong, Hong Kong, David Yao

We develop a model for manufacturers who are considering switching from existing technologies to clean technologies that generate fewer carbon emissions. Such a switch requires upfront fixed investment cost. For instance, in cement manufacturing, investment in Carbon Capture and Storage (CCS) is needed for cleaner production. Another example is the paper and pulp industry, where cleaner production technologies entail replacing the fossil-based heaters with heat pumps. Clearly, such investments, given their upfront nature, are risky, so risk-averse producers may not want to switch to the clean technologies. Our objective is to develop risk hedging strategies for such producers, by trading carbon emission permits in a secondary market, such as the EU Emissions Trading System (ETS) market, as well as other relevant financial assets that impact product demand (for example, stock index as a proxy of the general economy). We jointly optimize hedging strategy, capacity decision and technology choice to maximize Conditional Value at Risk (CVaR) objective. The solutions are completely characterized analytically. We find that risk hedging facilitates the clean technology adoption by allowing for larger upfront investment cost, compared to the production-only model.

3 - AI Algorithmic Pricing and Collusion

Yuze Li, Questrom School of Business, Boston University, Boston, MA, United States, Jussi Keppo, Gerry Tsoukalas, Nuo Yuan

The increased use of algorithmic pricing agents poses significant challenges for markets and regulatory oversight. This paper provides a theoretical and experimental investigation into algorithmic pricing agents based on Large Language Models (LLMs). We find that (1) LLM-based agents can autonomously achieve optimal pricing, and (2) they engage in pricing collusion without explicit communication. To investigate the underlying mechanisms, we develop a two-period stylized model that motivates an extensive simulation exploration into various factors influencing the rate of pricing collusion among LLM-based agents, including (1) the number of agents,

(2) heterogeneity in algorithmic and computational capabilities, (3) variability in the agents' information environment, and (4) the degree of platform competition. Our findings uncover the dynamics underpinning algorithmic collusion, providing unique insights into regulating LLM-based pricing agents and algorithmic market behavior.

4 - Costless Multi-Party Signaling with Buyback Contracts in Supply Chain Finance

Wenqiang Xiao, New York University, New York, NY, United States, Fangruo Chen, Guoming Lai, Chuanshuai Ru

We investigate an innovative financing scheme in which large suppliers use buyback clauses to help downstream resellers secure favorable loan terms. A key question is how buyback contracts can operate effectively when external lenders have less knowledge about a product's prospects than supply chain parties. We develop a multi-party signaling framework involving three parties: a supplier, a financially constrained reseller, and a lender, to examine the value of buyback contracts as an information-revelation instrument. In our model, while offering buyback terms can reduce loan interest rates by sharing default risk, it does not inherently improve supply chain performance under symmetric information. In contrast, under asymmetric information, the choice of buyback terms becomes critical. Sufficient differentiation between contracts—such as lower wholesale prices or higher buyback prices under the optimistic product prospect compared to the pessimistic one—combined with aligned procurement quantities, allows the supply chain to credibly signal the product's prospects to the lender in a costless way. The supplier's access to information proves pivotal in facilitating efficient signaling. Our findings underscore the powerful capacity of buyback contracts as both risk-sharing and signaling tools, providing practical insights for large suppliers on contract design and information management. These benefits are unattainable through standard wholesale-price contracts. Additionally, the multi-party signaling framework offers a generalizable solution methodology applicable to similar signaling scenarios, extending its relevance beyond supply chain finance.

TD44

Building B Level 3 B304

Food Supply Chain and Waste

Contributed Session

Contributed

Chair: Zahra Hossein Pour, University of Arkansas, 800 W Dickson St, Fayetteville, AR, 72701, United States

1 - Optimal Pricing and Date Labeling Strategies for New and near-Expiry Food Products: Implications for Food Waste Reduction

Shuxian Xiao, University of Northern Iowa, Cedar Falls, IA, United States, Charles Wang, Mike Wei

In this paper, we study three common retail practices for communicating expiration dates to consumers: (1) **Strategy DN**, which involves displaying and selling only new products; (2) **Strategy DB**, which involves displaying date labels and offering different prices for new and near-expiry products; and (3) **Strategy M**, which involves mixing new and near-expiry products while charging a uniform price. Our research is guided by the following questions: *How do consumers behave under these three approaches? What are the retailer's optimal prices, profits, and resulting food waste under each approach? Which strategy performs best in terms of financial outcomes and food waste reduction?*

We develop a Stackelberg game in which the retailer first determines its date-labeling and pricing strategy to maximize profit. Consumers then decide whether to purchase a new product, a near-expiry product, a product from the mixed batch, or not to purchase at all.

We analytically examine how these date-labeling strategies influence consumer purchasing behavior, retailer pricing decisions, profit outcomes, and food waste. Our analysis identifies optimal strategies and the conditions under which mixing new and near-expiry products can simultaneously reduce food waste and enhance retailer profitability. We also develop an infinite-horizon model in which the retailer replenishes the inventory of new products at the beginning of each period. The results and insights from the infinite-horizon model are qualitatively consistent with those from the single-period setting.

2 - Inventory Competition, Stockouts and Food Waste

Timothy Richards, Arizona State University, Mesa, AZ, United States, Stephen Hamilton, Elliot Rabinovich

The economic cost of lost or wasted food at the grocery-retail level is well documented. We analyze how much of this loss results from grocers competing in both price and inventory as these retailers overorder to avoid stockouts. We use grocery retail data to estimate a model of spatial demand, and competition in prices and inventories, among competing retailers to examine stockout costs and retail food loss. We find that prices and inventories are strategic complements in the sense that when retailers increase inventory levels to avoid stockouts, doing so softens price competition so prices rise accordingly in Bertrand-Nash equilibrium. Moreover, because we account for the reverse effect as well -- the feedback effect between prices and inventories -- we find that higher prices in turn lead to higher inventory levels. We find that, at the Bertrand-Nash equilibrium, retailers over-order relative to their baseline demand so that they waste some 13% of retail sales, on average, over multiple categories. We also find higher stockout costs than in the prior literature, and excess ordering of some 17% relative to noncompetitive-retailer levels. Retail competition in inventory appears to be an important driver of excess food stocks and an overlooked market failure contributing to overall food waste in the supply chain. This explains why retailers waste so much food, as holding excess inventories is profit-maximizing when there is a low-cost way of moving excess inventories out the back of the stores without incurring tipping fees.

3 - Profitability and Food Waste in Buffet Restaurants

Yuwen Hu, University of Michigan, Ann Arbor, MI, United States, Ekaterina Astashkina, Izak Duenyas

All-you-can-eat buffets are often criticized for generating excessive food waste. We build a stylized model of a profit-maximizing buffet operator facing cost-minimizing consumers with uncertain appetites. First, we examine how improvements in food accessibility (i.e., enhanced convenience in a buffet) affect consumer plate waste and the buffet's profitability. Contrary to conventional wisdom, we find that greater convenience can lead to lower plate waste and results in higher profits, as easier plate refills encourage diners to take smaller food portions. We then compare two common strategies buffet restaurants use to reduce plate waste: (i) imposing penalties for leftovers and (ii) educating consumers about the environmental and social implications of waste. We find that imposing an optimal penalty consistently boosts profits, whereas consumer education may sometimes lower profits, particularly in buffets with low unit production costs. Furthermore, optimal penalties outperform consumer education in terms of profitability, but education can be more effective in reducing plate waste in buffets with moderate unit production costs. Our findings remain robust even in environments where consumers do not perfectly learn their demands, as well as in the presence of insatiable or environmentally conscious consumers.

4 - Governance Strategies Targeting the 'Hidden Middle' to Reduce Food Loss and Health Risks in China's Food Supply Chain

Cangyu Jin, Zhejiang University, Hangzhou, China, People's Republic of

The untraceability challenges posed by the often overlooked "hidden middle"—small- and medium-scale entities within complex agricultural supply chains—exacerbate food losses and health risks. However, the impact of the "hidden middle" within the food supply chain, from the perspective of cross-regional food safety risk transfer, and the effectiveness of governance, remain uncertain. Leveraging over 300,000 aquatic product tests from 247 prefectures between 2015 and 2022 in China, we developed a data driven framework examines food loss and health risks associated with antibiotic residues in aquatic supply chains, and proposes optimization strategies within the framework of the existing food sampling schemes. First, we found that widespread contamination by antibiotic residues exceeding maximum residue limits (MRLs) throughout China by conditional probability model. Second, we employed the conditional probability model with dose-response relationships revealed an average food loss of 44,200 tons per province and health risks exceeding warning thresholds in six provinces. Third, Integer Programming (IP) was used to optimize the allocation of food safety monitoring resources for aquatic products along the supply chain, aiming at maximizing the probability of detecting contaminated products. The optimization leading to an average mitigation of 14,702 tons of food loss and a 77% reduction in health risks from antibiotic residues in the national aquatic supply chain—improvements of 6,951 tons and 17% compared to the baseline monitoring approach. Our findings highlight the importance of the "hidden middle" in food safety governance along complex agricultural supply chains, which plays a critical role in reducing food loss and health risks.

5 - A Joint Optimization Model for Storing and Transporting A Perishable Product in A Cold Chain

zahra Hossein Pour, University of Arkansas, Fayetteville, AR, United States, Haitao Liao

One of the most challenging tasks in the food industry is controlling product quality throughout the cold chain. In this research, we consider two key stages in the cold chain for a perishable product: storage and transportation. For a given demand with a specific delivery window, a joint optimization model for storing and transporting the product is developed considering a potential loss due to product decay during the two stages. The objective is to determine the optimal order quantity, the start time of transportation process, and the storage and transportation conditions—particularly temperature settings—to minimize the total cost while satisfying the customer demand. The total cost includes the storage cost, transportation cost, and the cost of wasted units. To address this problem, the life distributions of the product due to decay during both the storage and transportation are jointly considered. We employ the cumulative exposure model to formulate product quality deterioration over time and its dependency on temperature settings. The proposed joint optimization model is applied in an illustrative case study to demonstrate its practical applicability. This model can serve as a valuable tool for designing and operating food distribution relying on a cold chain, offering an effective approach for balancing food quality preservation, delivery time, and cost across the cold chain.

TD45

Building B Level 3 B305

Topics in Empirical Operations Management

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Caleb Kwon, The University of Texas At AUstin, Austin, United States

1 - Racial and Gender Biases in Customer Satisfaction Surveys: Causal Evidence from An American Restaurant Chain

Yong-Pin Zhou, University of Washington, Seattle, Seattle, WA, United States, Masoud Kamalahmadi, Qiuping Yu

We study racial and gender biases in customer satisfaction surveys using 1.5 million transactions and 260,000 customer surveys from a casual-dining restaurant chain in the US. We find that customer ratings are biased against racial minorities, and strikingly female servers despite them being the majority. We explore the mechanisms of discrimination and propose mitigation strategies.

2 - First Impressions Matter: Task Frictions and Retention on a Crowdfork Platform

Martin Gonzalez Cabello, University of California, Los Angeles, Los Angeles, CA, United States, Auyon Siddiq, Charles Corbett, Andrew Gordon

Crowdfork platforms are central to the data supply chains that power AI systems, particularly through large-scale data labeling. Yet the dynamics of worker engagement on these platforms remain poorly understood. We study fairness and worker engagement on a large online crowdfork platform. Leveraging a comprehensive dataset that records every task completed during a full calendar year, we focus on workers that recently joined the platform and analyze how their early experiences shape subsequent platform activity. We examine several dimensions of procedural fairness and how they influence workers' future engagement with the platform. We link submission records, study metadata, and dashboard data to reconstruct what each worker saw before choosing a task. Using this information, we assess how early exposure to mismatched or returned tasks affects long-term engagement. Our findings have direct implications for platform design and policy: improving early task experiences may promote retention and build trust among new workers.

3 - 2004--2024: Unpacking Two Decades of Food Date Labeling in the US and UK

Fiorentia Zoi Anglou, University of Texas at Austin, Austin, TX, United States, Robert Bray, Robert Sanders, Ioannis Stamatopoulos

Food date labels---e.g., "best by," "use by," "sell by"---are widely believed to influence consumer behavior and therefore contribute to food waste. Based on these beliefs, policymakers have taken action: California will soon implement date-label changes, a date-label-standardization bill is pending in Congress, and five of the top-10 UK retailers have switched from "use by" to "best by" labels. Our research provides the first large-scale, field study assessing the impact of date labels on consumer purchasing behavior. We combine data from 40,000-60,000 households across the US with state-level policy variation for dairy products' date labels in four states between 2005--2021. Using an event-study design and placebo inference, we find no evidence of change in consumer behavior in response to changes in date labels. These findings directly challenge prevailing assumptions among policymakers and regulators about the relationship between date labeling and food waste.

4 - The Digital Hustle: Can After-Hours Work Emails Shape Career Success

Priscilla Rodriguez, INSEAD, Fontainebleau, France, So Yeon Chun, Jurgen Mihm

Today's professionals have unprecedented access to digital communication tools that enhance connectivity and control over work-related tasks. However, these tools can be a double-edged sword: employees may be pressured to remain constantly available, creating an 'always on' work culture. In the past, clear boundaries limited the amount of work employees could do outside scheduled hours, but now these boundaries have blurred. Thus, employees may feel compelled to engage in after-hours work, defined as work beyond scheduled work hours. Yet a critical question remains: does this even pay off? This question has far-reaching implications not only for individual career trajectories but also for family dynamics, mental health, and broader workplace culture.

In this study, we empirically examine the impact of after-hours work on performance evaluations using proprietary, anonymized data from over 3,000 desk job employees at a Fortune 500 company. This dataset integrates granular digital traces of work behavior, including email and meeting records with timestamps and participant details, individual performance evaluations, and HR metadata on role, seniority, and demographics. To mitigate potential endogeneity, we use a novel instrumental variables approach that leverage both formal reporting relationships and informal collaboration networks. We investigate whether and how after-hours work influences evaluations, propose a potential mechanism, and provide supporting empirical evidence. Our findings offer insights into how after-hours work patterns may influence career outcomes while highlighting that the effects vary across individuals.

TD46

Building B Level 3 B306

Simulation and Modeling

Invited Session

Decision Analysis Society

Chair: Kamelia Sepanloo, Purdue University, West Lafayette, IN, United States

1 - Multimodal Evaluation of Nurse Trainees in Extended Reality

Kamelia Sepanloo, Purdue University, West Lafayette, IN, United States, Young-Jun Son, Janine Hinton

This presentation examines the use of eye-tracking and physiological response analysis to assess cognitive load and stress in nursing trainees during simulation-based education in extended reality environments. The study utilizes a custom-designed training system that operates in both mixed reality (MR), using the HoloLens 2 headset, and virtual reality (VR), using the Meta Quest Pro. Nurses interact with digital patient and digital medical equipment in MR, and within a fully immersive digital environment in VR. Throughout the simulation, participants' eye movements are continuously recorded to capture metrics such as fixation duration, saccade frequency, and gaze distribution, offering insight into visual attention and task engagement. In parallel, physiological data are collected using the Empatica Embrace plus wristband, capturing heart rate, electrodermal activity (EDA), and skin temperature. These measures serve as indicators of stress and arousal during high-stakes clinical scenarios, including moments when the patient's condition deteriorates due to hypotension and hypoxia. The training simulation is segmented into six time-based stages, allowing for

detailed temporal analysis of both gaze behavior and physiological responses as the scenario evolves. Preliminary findings indicate notable differences in cognitive load and stress responses between MR and VR environments, with implications for the design of adaptive training systems. By combining eye-tracking and physiological data, this research provides a multidimensional understanding of trainee experience and offers a pathway toward more responsive, personalized simulation-based education. Ultimately, these insights aim to enhance the preparedness of healthcare professionals by optimizing learning environments based on real-time cognitive and emotional feedback.

2 - Real-Time Digital Twin Development and Augmented Environment Generation for Multi-Agent Robotic Systems

Seonho Woo, Purdue University, West Lafayette, IN, United States, Seung Ho Woo, Young-Jun Son

This study focuses on the development of a real-time bidirectional digital twin framework utilizing the NVIDIA Isaac Sim platform. Current robotic systems struggle with effective decision-making in dynamic environments due to limited sensory input, data inaccessibility from occluded regions, and resource constraints. Our study addresses these challenges through the proposed interaction framework, leveraging a generative model, where physical and digital twins continuously exchange information to enhance situational awareness and generate environmental knowledge.

Our system constructs high-fidelity virtual environments using multimodal sensor data from LiDAR, cameras, and IMU sensors mounted on robotic platforms. The bidirectional communication enables real-time experiment planning where the digital twin can simulate navigation scenarios and path optimization strategies while the physical twin validates and refines these predictions through actual execution. A key innovation involves developing generative AI models that create comprehensive environmental representations using limited real-time data without relying on extensive historical datasets or pre-defined knowledge bases.

The current implementation focuses on perfecting single-agent digital twin interactions of UGV, with planned expansion to multi-agent systems that demonstrate collaborative decision-making through distributed sensing and coordinated task execution. This approach optimizes resource utilization and maintains data integrity through privacy-preserving frameworks. The system's scalability positions it for large-scale digital twin modeling, which would be a practical approach for existing challenges in manufacturing facility and warehouse automation that can expand to support efficient decision-making in dynamically changing environments.

3 - Properties of Two-Stage Stochastic Multi-Objective Linear Programs

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

We study a two-stage stochastic multi-objective linear program (TSSMOLP) under risk neutral formulation, which is a natural generalization of the well-studied two-stage stochastic linear program (TSSLP). The proposed TSSMOLP formulation allows modelers to specify multiple objectives in each stage. The second-stage recourse decision is modeled by a random multi-objective linear program (MOLP) whose solution maps to a random second-stage nondominated set. The TSSMOLP then comprises the objective, which is the Minkowski sum of a linear term plus the expected value of the second-stage nondominated set, and the constraints, which are linear. We prove properties of TSSMOLPs, which are stochastic set optimization problems, and the multifunctions that arise therein. To aid theoretical analysis, we propose nondominance-equivalent reformulations to the original problem and show that (i) the global Pareto set of a TSSMOLP is cone-convex, (ii) solving a parameterized TSSLP produces a weakly efficient point for the TSSMOLP, and (iii) a TSSMOLP in which the uncertainty is governed by an atomic probability measure with finite support can be reformulated as an MOLP. Leveraging the above properties we demonstrate that solving a TSSMOLP under a sample average approximation framework is equivalent to solving a large scale MOLP, and validate our results through a numerical example.

TD47

Building B Level 3 B307

Optimization Topics 1

Contributed Session

Contributed

Chair: Peng Ren, George Mason University, Fairfax, VA, 22030, United States

1 - Reducing Contextual Stochastic Bilevel Optimization via Structured Function Approximation

Maxime Bouscary, MIT, Cambridge, MA, United States, Jiawei Zhang, Saurabh Amin

Contextual Stochastic Bilevel Optimization (CSBO) extends standard stochastic bilevel optimization (SBO) by incorporating context-dependent lower-level problems, as in hyperparameter tuning and inverse optimization. This added structure introduces significant computational challenges: solving CSBO requires solving an infinite number of lower-level problems - one for each context realization - and existing approaches either suffer from high sample complexity or rely on impractical conditional sampling oracles. We propose a reduction framework that approximates the lower-level solutions using expressive basis functions, thereby decoupling the lower-level dependence on context and transforming CSBO into a standard SBO problem solvable using only joint samples from the context and noise distribution. Under mild assumptions, we show this reduction preserves hypergradient accuracy and yields an ϵ -stationary solution to CSBO. We relate the sample complexity of the reduced problem to simple metrics of the basis and show that using Chebyshev polynomials leads to a near-optimal complexity of $\tilde{O}(\epsilon^{-3})$, matching the best-known rates for standard SBO. Empirical results on hyperparameter and inverse optimization tasks demonstrate that our approach outperforms CSBO baselines in convergence, sample efficiency, and memory usage, especially in settings without conditional sampling access.

2 - Admission Control for A Single Server Waiting Time Process in Heavy Traffic

Bowen Xie, The University of Akron, Akron, OH, United States, Haoyu Yin

We investigate a single-server queue control problem (QCP) in a heavy-traffic regime, extending the framework from Lee and Weerasinghe (2011). The state process represents the offered waiting time, the customer arrival has a state-dependent intensity, and the customers' service and patience times are i.i.d generally distributed. We introduce an infinite-horizon discounted cost functional consisting of a control cost and a penalty for idleness cost. Our primary goal is to tackle the QCP, taking into account a non-trivial polynomial decaying cost function resulting from the control mechanisms in the waiting time. Under mild assumptions, the heavy traffic limit of the QCP yields a control problem described by a diffusion process, referred to as a diffusion control problem (DCP). We find the optimal control of the associated DCP by incorporating the Legendre-Fenchel transform and a formal Hamilton-Jacobi-Bellman (HJB) equation. Then, we "translate" this optimal strategy to the QCP, of which we obtain the asymptotic optimality. Beyond theoretical analysis, we explore the application of the REINFORCE algorithm, a reinforcement learning (RL) approach, to solve stochastic control problems, motivated by recent advances in the field. We compare theoretical insights with data-driven algorithmic solutions, highlighting the advantages and limitations of both simulation-based and analytical approaches.

3 - Ten Years of CVXPY

William Zhang, Polytechnique Montreal, Montreal, QC, Canada, Steven Diamond

The CVXPY project has been under development for over ten years. The initial motivation was to reproduce in Python the functionality of the CVX MATLAB package. However, early CVXPY design decisions diverged from CVX, with a greater emphasis on modular problem construction, expression trees, and parameterized problems. We explore how these initial design decisions led to later research on object oriented optimization, matrix-free optimization, and differentiable optimization. We conclude with an overview of upcoming CVXPY features.

4 - Advances on Multi-Parametric Dynamic Optimization Problems

Lida Lamakani, Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, United States, Efstratios Pistikopoulos

The increasing complexity of engineering systems demands advanced control strategies. Model Predictive Control (MPC) has emerged as an advanced optimization-based control that uses system models to predict future behavior and manage constraints. Traditional MPC solves the optimization problem at each sampling step, which can be

computationally demanding. Multiparametric programming offers a more efficient alternative by solving the optimization problem offline and partitioning the state space into regions, each with an explicit control law, enabling fast, real-time decisions without online computation.

Most multiparametric MPC approaches are in discrete time, which may cause inaccuracies in systems with fast dynamics or long sampling intervals. Continuous-time formulations preserve the system's natural behavior and ensure constraints are satisfied throughout the horizon. In this work, both linear and nonlinear dynamic MPC problems are solved by combining the Hamiltonian function and Pontryagin's Maximum Principle with the multiparametric framework instead of the KKT conditions, resulting in fewer critical regions and the identification of switching points, when system behavior changes due to constraint activity, which can additionally accelerate online implementation.

For nonlinear systems, several strategies are explored to solve nonlinear ODEs that result from the above optimality conditions. First, the Koopman operator lifts nonlinear space into linear space but may struggle with small powers. Second, a regression-based method uses the structure of linear ODEs to approximate expressions for solutions obtained from numerically solved nonlinear systems, though sensitivity to time and nonlinearity remains a limitation. Finally, neural networks are used to capture nonlinear behavior while enabling interpolation and extrapolation.

5 - Solving Latency Location-Routing Problem Using Dynamic Programming Method

Peng Ren, George Mason University, Fairfax, VA, United States, Rajesh Ganesan

The Latency Location-Routing Problem (LLRP) seeks to minimize cumulative customer waiting time by determining both which depots to open and how vehicles should be routed to minimize total latency while fulfilling demand. This problem is especially critical in situations such as post-disaster logistics, where timely delivery and responsiveness are essential. Yet the existing literature remains limited: most models ignore real-world uncertainty in travel times and demand, and few approaches scale well or support adaptive decision-making. Challenges arise in simultaneously accounting for evolving system state including depot configurations, vehicle positions, remaining vehicle capacities, and dynamically changing demand over time.

This work introduces a dynamic programming (DP) formulation of LLRP that addresses these challenges. The model captures the system state as a combination of depot status, vehicle locations, residual capacities, and stochastic demand, and models their evolution through probabilistic transitions. Dynamic programming is well suited for such problems due to its ability to handle cumulative cost objectives, sequential routing and allocation decisions, and adaptation as uncertainty unfolds. While exact DP may be computationally intensive, this framework enables the development of scalable approximate dynamic programming or learning-based strategies tailored for time-sensitive, uncertainty-rich environments.

A small case study demonstrates how the model structure accommodates evolving network conditions and demand realization. This work contributes a new formulation perspective for LLRP, enabling more adaptive and interpretable decision-making and addressing key modeling gaps in the literature.

TD48

Building B Level 3 B308

Behavioral Decision Analysis II

Invited Session

Decision Analysis Society

Chair: Pelin Gülüm Taş, Delft University of Technology, Delft, 2518SE, Netherlands

1 - Cognitive and Motivational Components of Normalcy Bias: a Signal Detection Approach

Richard John, University of Southern California, Los Angeles, CA, United States, Robin Dillon, Kevin Kapadia, Nicholas Scurich

Officers at the United States Coast Guard (USCG) Search and Rescue (SAR) make a variety of high-stakes decisions under conditions of uncertainty and imperfect information. A voluminous body of psychological research has documented that many decisions in these situations are subject to various biases that can lead to poor outcomes. For the USCG, these can be catastrophic outcomes, such as when the USCG fails to initiate a search and rescue or when a search and rescue mission is unnecessarily initiated, putting officers' well-being at risk and incurring significant financial expenses. This project seeks to identify potential cognitive and motivational biases that might impact USCG decision making and determine how these biases affect decisions that result in normalcy bias, wherein warning signals and threats are disbelieved. We utilize a normative Signal Detection Theory (SDT) framework, which partitions decision accuracy into (1) the ability to discriminate signals from noise and (2) the selection of a decision threshold. This framework illustrates how cognitive and motivational biases can affect both components in SAR decisions made by the USCG. We report empirical behavioral research to understand deviations from the normative SDT model, including Normalcy Bias.

2 - Gain-Loss Bias Within Multi-Attribute Value Theory

Geqie Sun, Delft University of Technology, Delft, Netherlands

Human judgments are essential inputs in decision analysis methods, yet behavioral decision research has shown that such judgments are susceptible to various cognitive biases. When these biases affect the inputs, they can undermine the quality of the methods and their results. This study investigates gain-loss bias, a well-documented cognitive bias, within the context of Multi-Attribute Value Theory (MAVT). We focus on two sources of potential bias: (i) attribute framing: We examine whether framing the same attribute as a gain versus as a loss influences the elicited attribute-specific value functions and weights (scaling constants). (ii) tradeoff framing: In MAVT's tradeoff procedure for weight elicitation, decision-makers are asked how much of one attribute they are willing to give up or gain in exchange for some change in another attribute. Though the alternatives being compared are logically equivalent, framing the tradeoff as a gain (adjusting from the worst level) or a loss (adjusting from the best level) can influence the perceived value of the attributes due to gain-loss bias and distort the attribute-specific value functions and weights. To test these hypotheses, we designed a questionnaire based on the MAVT methodology and conducted an experiment involving 283 participants in a decision problem. The results show that gain-loss bias significantly affects both the attribute-specific value functions and weights, and consequently, the decision results.

3 - Time Preference in Multi-Attribute Decision Making: An Experimental Study

Pelin Gülüm Taş, Delft University of Technology, Delft, Netherlands

Research on time preference has predominantly employed descriptive approaches, focusing on intertemporal trade-offs within single-objective decision-making frameworks. However, real-world problems are often multi-objective, requiring trade-offs not only over time but also between conflicting objectives. This complexity calls for prescriptive approaches that move beyond description and support decision-making in the face of intertwined trade-offs. Addressing this gap, the present study investigates time preferences within a multi-attribute decision-making (MADM) context using the Multi-Attribute Value Theory (MAVT) framework. First, we theorize the potential impact of time discounting on attribute-specific value functions (AVF) and scaling constants (weights) (two main components of MAVT), formulating four hypotheses. Then, three controlled experiments are conducted around an intertemporal MADM problem involving two attributes. A total of 279 participants took part in the experiments, each completing two scenarios in which the timing (present vs. future) of outcomes varied in relation to specific attributes. The results indicate that both the scaling constants and the perceived AVFs vary significantly with delays. Furthermore, the underlying drivers of these changes depend on the specific trade-off relationships involved. While delaying a single outcome typically leads to discounting, our findings indicate that postponing multiple outcomes in multi-objective contexts does not necessarily result in the same behavior. The presence of multiple objectives challenges conventional assumptions about delay discounting. These insights are particularly relevant for researchers and practitioners in economics, psychology, health, and public policy who deal with complex, multi-dimensional intertemporal decisions.

4 - Fairness Concerns in 3-Tier Contests

HUA CHEN, University of Georgia, Athens, GA, United States, Kevin Chung

Consider a contest conducted in a **3-tier** organization where the owner provides a winning prize for branches to compete and the manager of each branch offers a reward to motivate his agents. The standard economic theory predicts that the information on total winning prize or the reward offered to the opponent agents would have no impact on the effort decisions by the focal agent. Using an incentive-aligned lab experiment, contrary to the predictions, the authors find that: 1) the agents adjust their effort based on their knowledge about the total winning prize or the reward offered to the opponent agents; 2) the managers modify their reward upward (downward) when either (both) of the information is revealed to the agents. These behavioral regularities are explained by a model that incorporates the fairness concerns (vertical and horizontal) and the relative saliency of the two into the agent's utility function.

TD49

Building B Level 3 B309

Novel Applications in Empirical Research

Invited Session

MSOM: Technology, Innovation, and Entrepreneurship

Chair: Serguei Netessine, The Wharton School, Philadelphia, PA, United States

1 - Adaptive Sampling for Learning Rankings from Few Evaluators

Motti Goldberger, Yale School of Management, New Haven, CT, United States, Nils Rudi

Organizations that must choose among multiple alternatives—whether hiring job candidates, prioritizing R&D projects, or deciding which athletes to draft—often rely on input from evaluators. The set of evaluators may be small, either due to limited availability of domain experts or confidentiality constraints. Furthermore, each evaluator's capacity to provide input may be limited: their time is costly, and their accuracy may degrade as the scope or complexity of the evaluation increases. In such settings, the choice of whom to query and what input to collect becomes important.

We consider a setting where an organization aims to learn a complete ranking of a set of alternatives. Although collecting full rankings from evaluators would be ideal, asking them to rank many items can be cognitively demanding and reduce the reliability of their input. Instead, we focus on simpler elicitation formats—pairwise comparisons and ordinal ratings—which are easier for evaluators to provide while still informative for learning the overall ranking. We allow for the possibility that evaluators' beliefs may be biased and their responses may be noisy.

To model this, we adopt a hierarchical Bayesian framework in which both the true ranking and each evaluator's belief are treated as random. The aim is to sequentially collect inputs that minimize posterior uncertainty about the true ranking. To that end, we develop an adaptive sampling algorithm that, at each step, chooses both which evaluator to query and which comparison to elicit, making the most efficient use of limited evaluation opportunities.

2 - Dynamic Staffing for Harvesting Under Climate and Labor Pool Uncertainties

Sanchita Das, University of Washington, Seattle, WA, United States, Masha Shunko, Leonard Boussieux

Eastern Washington's tree fruit industry is grappling with the dual challenges of accelerating climate change, leading to more frequent heatwaves than ever, and evolving labor regulations that are reshaping the workforce. This research addresses the critical problem of optimizing harvest labor management strategies for this vulnerable sector by collaborating extensively with the regional agricultural community. The study develops a robust model using a combination of stochastic dynamic programming and distributionally robust optimization to account for uncertainties in ripe fruit forecasts and seasonal weather patterns. Key drivers of profitability, such as picking timeliness and adaptive hiring policies, are utilized to develop the model. Results indicate that the proposed methodology significantly increases profitability compared to baseline staffing models, generating a positive net present value over a multi-year planning horizon while enhancing resilience to these growing uncertainties. The findings underscore the urgent need for a coordinated response from government, farmers, labor platforms, and other stakeholders to mitigate the economic threats posed by global warming and changing labor landscapes. This research provides a fundamental model to quantify and optimize the impact of various initiatives on harvest labor management, offering heuristics to guide farmers in maximizing profits amidst climate and labor-related uncertainties.

3 - More Money, More Motivation? a Field Experiment on the Impact of Bonus on Service Performance in a Multitasking Setting

Tom Tan, Southern Methodist University, Dallas, TX, United States, Jia Li

In service settings, workers often handle multiple tasks of varying difficulty. While financial incentives are commonly used to motivate performance on challenging tasks, they may unintentionally distort behavior by diverting effort from other tasks. We study this issue through a six-week field experiment with the after-sales department of a Fortune Global 100 online retailer. Gig agents on a live-chat platform were randomly assigned to a treatment group, which received real-time notifications for difficult cases (flagged by an AI model) indicating a bonus opportunity, or to a control group without such notifications.

We analyze customer satisfaction scores and redial rates as performance measures across four case difficulty levels. On average, financial incentives improved satisfaction scores. However, we find heterogeneity by task difficulty: for Level 1 (most difficult) cases, customer satisfaction declined in the treatment group, while satisfaction improved for easier, non-incentivized cases. These patterns suggest that agents strategically reallocate effort based on perceived attainability of rewards. When the likelihood of earning a bonus on very difficult cases appears low, agents redirect effort toward easier cases with more achievable outcomes.

Our study highlights how incentives may trigger unintended motivational spillovers in multitasking environments. Workers may reduce effort on the very tasks the incentives were meant to improve. We contribute to the literature by showing that in such settings, bonus design must consider strategic effort allocation. We also suggest that combining monetary and non-monetary interventions—like real-time AI support—may better address effort challenges on the most difficult tasks.

4 - From Information Aggregators to Transaction Facilitators: An Empirical Investigation of the Impact of the Platform Payment Integration

Serguei Netessine, The Wharton School, Philadelphia, PA, United States, Shi Zhenming, Vivek Choudhary

Over the last decade, digital platforms have grown significantly, broadly operating under two distinct models from the perspective of where the final transactions occur. Some platforms function as “information aggregators”, solely matching buyers with sellers, while transactions occur outside the platform. Others serve as “transaction facilitators”, integrating payment systems that enable users to complete transactions directly within the platform. What is the impact of payment integration on various stakeholders when a platform transitions from an information aggregator to a transaction facilitator? We examine this evolution in the novel context of trucking platforms. Using a quasi-experimental study of a major Chinese freight logistics platform that introduced transaction capabilities, we analyze over 2.6 million transportation records through a Difference-in-Differences (DID) approach to understand how this transition affects platform stakeholders. Our results show that integrating a transaction system improved operational efficiency by reducing order acceptance times and delivery costs. Weekly driver earnings increased by ~15%, primarily because the platform's transaction system likely builds trust between drivers and unfamiliar shippers, leading to higher order acceptance rates and better resource utilization. However, earnings per kilometer per ton decreased by ~3.4%, suggesting drivers needed to handle larger workloads to maintain their income levels.

TD50

Building B Level 3 B310

Causal Inference and Machine Learning in Operations Management

Invited Session

MSOM: Service Operations

Chair: Ruomeng Cui, Emory University, Amazon, Decatur, GA, United States

Co-Chair: Zhikun Lu, NYU Shanghai, Shanghai, China, People's Republic of

1 - Synthesizing Evidence: Data-Pooling for Treatment Selection in Online Experiments

Renyu Zhang, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Zhenkang Peng, Chengzhang Li, Ying Rong

Randomized experiments are the gold standard for causal inference but face significant challenges in business applications, including limited traffic allocation, the need for heterogeneous treatment effect estimation, and the complexity of managing overlapping experiments. These factors lead to high variability in treatment effect estimates, making data-driven policy roll-out difficult. To address these issues, we introduce the data-pooling treatment roll-out (DTR) framework, which enhances policy roll-out by pooling data across experiments rather than focusing narrowly on individual ones. DTR can effectively accommodate both overlapping and non-overlapping traffic scenarios, regardless of linear or nonlinear model specifications. We demonstrate the framework's robustness through a three-pronged validation: (a) theoretical analysis shows that DTR surpasses the traditional difference-in-mean method under non-overlapping experiments, particularly when the number of experiments is large; (b) synthetic simulations confirm its adaptability in complex scenarios with overlapping traffic, rich features and nonlinear specifications; and (c) empirical applications to experimental data from online platforms showcase its strength in group-based policy roll-out. By reducing estimation variability to improve decision-making accuracy and efficiency, DTR provides a scalable, practical solution for online platforms to better leverage their experimental data in today's increasingly complex business environments.

2 - Incentives In Online Gaming: Optimal Policy Design with Dynamic Causal Machine Learning

Zhikun Lu, Emory University, Decatur, GA, United States, Ruomeng Cui

The video game industry has experienced remarkable growth over the past few decades, with online gaming emerging as a key driver of its recent expansion. Fueled by an engagement-centric business model, game developers are increasingly relying on data-driven insights to optimize player engagement, particularly through incentive programs that reward players with virtual items. This paper explores the design of such reward systems, focusing on the trade-off between offering in-game currency (which provides player autonomy) and algorithmically assigned specific items (which represent system control). We examine this trade-off in the context of Honor of Kings, a highly popular mobile multiplayer online game developed by Tencent. Using real-time player data, we find that in-game currency outperforms specific virtual items in some cases, but not universally, highlighting the importance of aligning rewards with players' understanding and experience. Our findings also reveal significant heterogeneity in player responses, suggesting that personalized reward strategies are crucial for maximizing engagement. To address this gap, we propose a data-driven framework for optimizing sequential treatment assignments in incentive programs under resource constraints, integrating dynamic causal machine learning with dynamic programming. It maximizes long-term user retention by dynamically adjusting reward allocations based on real-time player features. Our approach demonstrates substantial improvements in user retention and provides valuable insights that can extend beyond the gaming industry.

3 - No More Hidden Fees: The Impact of California's Honest Pricing Law on Vacation Rentals

Serkan Mehder, University of Colorado Boulder, Boulder, CO, United States, Maxime Cohen, Armin Khamoshi, Ovunc Yilmaz

Hidden fees are a common and controversial pricing practice in the short-term lodging industry. In July 2024, California implemented the Honest Pricing Law, requiring platforms to display all-in prices upfront. To study the impact of this regulation on host and customer behavior, we collaborate with the Expedia Group—the parent company of Vrbo—and analyze U.S.-wide click-level data covering more than one million vacation rental properties. Using a quasi-experimental approach, we examine how the law influenced consumer booking behavior and host pricing decisions.

4 - Generative AI and Content Homogenization: the Case of Digital Marketing

S. Alex Yang, London Business School, London, United Kingdom, Chaoran Liu, Tong Wang

Traditionally, entrepreneurs and small businesses, constrained by limited resources, struggle to create high-quality marketing content, which limits consumer engagement and hampers growth. Generative artificial intelligence (Gen AI) presents an affordable, powerful tool to overcome this challenge. However, the widespread adoption of Gen AI may have unintended consequences. Specifically, content generated by popular large language models may lead to increased homogenization, reducing consumer engagement and diluting brand uniqueness over time. To examine this possibility, we study the restaurant industry—a highly fragmented sector where 70% of the establishments are owned by independent entrepreneurs. We leverage the country-wise ChatGPT ban in Italy during April 2023 as a natural experiment to examine how access to Gen AI affects the content homogenization, consumer engagement, and volume of social media marketing content created by small business owners. Our study shows that the treatment group experienced significant decreases in different similarity measures during the ban, suggesting that access to ChatGPT leads to content homogenization. We also find the ban has a non-negligible effect on consumer engagement. By better underscoring an unintended consequence of Gen AI adoption, this research aims to provide valuable insights to entrepreneurs and small businesses on how to better harness Gen AI's benefits while mitigating potential drawbacks.

5 - A Deep-DiD Method to Estimate Heterogeneous Treatment Effects: Application to Content Creator Selection

Yan Cheng, Shanghai University of Finance and Economics, Shanghai, China, People's Republic of, Jingbo Wang, Xinyu Cao, Zuojun Max Shen, Yuhui Zhang

In this paper, we propose a Deep-DiD method that incorporates two deep neural networks in a difference-in-difference (DiD) framework to estimate heterogeneous treatment effects (HTEs). The dual-network architecture contains one neural network modeling HTEs as a nonparametric function of pre-treatment features, and another neural network capturing individual and time fixed effects. Through a series of simulations, we show that our method can uncover the true HTEs with high accuracy under various settings and demonstrates more robust estimation performance compared to existing methods like linear models and random forests. We apply this method to an empirical setting where a large video-sharing platform introduced a "Creator Signing Program" aimed at signing creators and motivating them to generate more high-quality video content. Leveraging a matched dataset of signed and unsigned creators, we employ our Deep-DiD method to estimate the HTEs of the signing program. Our method can help the platform optimize creator selection by identifying creators with the highest estimated treatment effects. Through out-of-sample tests, we show that creators selected by the Deep-DiD method experienced substantially larger actual performance jumps than those selected by the platform. Creator selection based on the Deep-DiD method also consistently outperforms that based on linear models.

6 - The Impact of Recommender Systems on Content Consumption and Production: Evidence from Field Experiments and Structural Modeling

Zhiyu Zeng, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Zhiqi Zhang, Tat Chan, Dennis Zhang

Online content-sharing platforms such as Facebook and TikTok, where users act as both content creators and consumers, have become integral to daily life, relying on complex algorithms to recommend user-generated content (UGC). While prior research and industry efforts have primarily focused on designing recommender systems to encourage users to consume more content, the impact of the content recommended to users on their own content production remains understudied. To address this gap, we conducted a large-scale randomized field experiment on a major video-sharing platform. Users were randomly assigned to either a treatment or control condition, and we manipulated content recommendations by excluding a subset of highly popular creators from being recommended to treated users. This reduced recommendation quality of treated users compared to control counterparts. Our results indicate that this decline led to a 1.34% decrease in video-watching time but a 2.71% increase in videos uploaded by treated users. These findings reveal a trade-off in recommender system design: while higher recommendation quality boosts consumption, it may suppress production. To optimize recommendations, we developed a structural model in which users' choices between consuming and producing content are inversely affected by recommendation quality. Counterfactual analyses suggest that slightly lowering recommendation quality—rather than prioritizing the most popular creators—can maximize the overall value of both production and consumption for platforms. This highlights the need to balance content consumption and production when designing recommender systems on these social media platforms.

TD51

Building B Level 3 B311

Perspectives on Food Security Management

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Ali Aouad, MIT, Cambridge, MA, United States

Co-Chair: Alp Sungu, Wharton, Philadelphia, PA, United States

1 - Long-Term Policy Prediction with Causal Graphs: Insights from Healthy Grocery Shopping

Nur Kaynar, Cornell University, Ithaca, NY, United States, Dmitry Mitrofanov

Policymakers often implement short-term interventions (e.g., temporary healthy food subsidies) to achieve long-term outcomes (e.g., sustained healthy eating habits). However, the delay in observing long-term effects complicates the immediate evaluation of policy effectiveness. To address this, we introduce a causal structure learning algorithm that identifies the causal relations among policy intervention, short-term variables, and long-term outcomes.

Our algorithm combines short-term experimental and long-term observational data, producing a graph that specifies causal links through directed edges. This approach identifies (i) which short-term variables are influenced by policy changes and (ii) which of these variables subsequently impact the long-term outcome. By mapping these pathways, we provide a framework for selecting short-term surrogates as early indicators of long-term effects. We then derive a closed-form expression to estimate long-term treatment effects without waiting for the long-term data to materialize.

We apply the framework to data from the U.S. Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) to assess the lasting effects of temporary healthy food subsidies. Our findings indicate that targeting subsidies toward younger consumers in urban areas with historically low healthy product consumption leads to long-term dietary improvements. Identifying effective short-term surrogates reduces the need for prolonged monitoring, minimizing costs and delays in policy evaluation. Overall, our paper provides policymakers with valuable insights and practical tools to enhance public healthy product subsidy strategies. By allowing them to evaluate long-term impacts through short-term randomized controlled trials, the framework we present helps refine policy design and optimize effectiveness.

2 - Food Safety Risk Propagation in the Agri-Food Supply Chain

Jane Wang, Massachusetts Institute of Technology, Cambridge, MA, United States, Retsef Levi, El Ghali Zerhouni, Yanchong Zheng

Food adulteration poses a serious global public health threat, particularly in countries with complex and fragmented agri-food supply chains (SCs) such as China. These risks can emerge at various points along the SC and propagate downstream to consumers. China's current food inspection strategy primarily targets retail stores, with the intent of encouraging safer sourcing from upstream suppliers. This study examines whether such downstream-focused testing effectively mitigates the propagation of adulteration risks. Using a self-constructed database of over 12 million food safety test records, it builds a panel dataset to quantify adulteration risks across multiple SC locations over time. Then, it develops a causal inference assessment of the product- and adulterant-level propagation effect of risks in wholesale markets and wet markets (WSMs/WMs) and manufacturers on risks in more downstream SC locations such as retail stores and restaurants. It provides strong evidence that adulteration risks propagate downstream, and existing testing strategies fail to mitigate this propagation. To design effective food inspection strategies, regulators should consider the introduction points and propagation paths of different adulteration risks. For instance, pesticide and veterinary drug residues—mainly introduced at farms—should be tested at WSMs/WMs, where fresh products are aggregated before reaching consumers. In contrast, food additives in processed products are best intercepted at the manufacturer level. This study highlights the limitations of existing testing strategies and provides a roadmap for more systematic risk-based food inspection policies within agri-food SCs.

3 - Systemic Analysis of U.S. Household Exposure to Food Ingredients and Nutritional Components

Cen hao Zhu, Massachusetts Institute of Technology, Cambridge, MA, United States

Access and consumption of “healthy foods” are increasingly recognized for their impact on human health and specifically on the prevalence of obesity and related chronic diseases. There are several public health policies and regulatory activities (the Healthy, Hunger-Free Kids Act of 2010, and MyPlate Plan by USDA, etc.) that aim at increasing access to healthy food, as well as ensuring that food products have appropriate nutrients and do not contain hazardous or unhealthy ingredients. However, many of the existing public health policies and much of the related research do not rely on sufficiently comprehensive and rigorous data. In this study, we employ advanced text analytics techniques to create the first large-scale dataset of branded food products sold in the U.S., delineating their corresponding ingredients and nutritional components, by combining multiple structured and unstructured data sources. Utilizing this data, we empirically characterize the prevalence of different food ingredients within specified product categories; the nutritional characteristics of typical household diets based on food products purchased; the overall exposure of households to different food ingredients; temporal changes over time; and differences among households across different regions. The resulting systematic understanding of household exposure to food ingredients and nutritional components will offer an essential scientific foundation for related public policy designs.

4 - Optimal Subsidy Design: Application to Food Assistance Programs

Zhicon Hu, INSEAD, Singapore, Singapore, Kamalini Ramdas, Ali Aouad, Alp Sungu

In the design of subsidy programs, a significant challenge lies in balancing policy efficiency with the flexibility of individual choice. Traditional paternalistic approaches often prioritize optimal social outcomes, such as promoting the consumption of nutritious but less-popular foods, potentially at the cost of individual preferences. This can lead to low take-up of the subsidy or low compliance, undermining the policy's effectiveness. Conversely, flexible subsidies, such as cash transfers, may result in suboptimal impact on nutrition. This study introduces the *optimal subsidy design problem*, which embeds a probabilistic choice model to represent the beneficiaries' decision to avail of an offered subsidy. Our formulation can optimize over various subsidy mechanisms, including in-kind bundles, assortments, vouchers, or cash-plus-in-kind mixes. After calibrating our model to data from a randomized controlled trial (RCT) on staple-food subsidies in India and synthetic data from AI simulations, we run counterfactuals to identify optimal subsidy designs. The analysis shows that pairing an in-kind staple subsidy with a modest cash top-up can outperform both pure in-kind and pure cash schemes on social outcomes and beneficiaries' choices.

5 - Improving Farming Practices Through LLM-Powered Digital Extension Services: a Randomized Controlled Trial with 2,000 Smallholder Rice Farmers

Canberk Ucel, ESSEC Business School, Paris, France, Ali Aouad, Alp Sungu, Niyazi Taneri, El Ghali Zerhouni

Low adoption and ineffective implementation of agricultural technologies continue to undermine productivity and sustainability in smallholder farming systems. This study presents one of the first large-scale randomized controlled trials (RCTs) testing the use of Large Language Model (LLM)-based AI advisory bots to address these challenges. We examine whether delivering contextualized, on-demand agricultural guidance via common messaging platforms on farmers' smartphones can improve farm management practices and the adoption of Direct-Seeded Rice (DSR) technology—a practice with demonstrated environmental and economic benefits. Our ongoing experiment targets smallholder rice farmers across five major rice-growing regions in the Philippines and India. Participants are randomly assigned to a control group or one of two treatment arms: access to an AI advisor bot, or access with additional incentives designed to encourage initial engagement. The advisory bot integrates technical content from Bayer CropScience and public sources and is tailored to local farming contexts. We collect detailed survey and interaction data throughout the growing season and evaluate technology adoption, farmer engagement, productivity, and AI-human interaction patterns. Findings will inform policy and industry efforts aimed at improving agricultural productivity, climate resilience, and the cost-effectiveness of extension services in developing countries. The study advances research at the intersection of agricultural development, operations management, and AI deployment in low-resource environments.

TD52

Building B Level 3 B312

Empirical Research in Socially Responsible Operations

Invited Session

MSOM: Sustainable Operations

Chair: Somya Singhvi, USC Marshall School of Business, Los Angeles, CA, United States

1 - The Ratings Trap: Prior Ratings, Gender and Instructor Evaluations

Kamalini Ramdas, London Business School, London, United Kingdom, Monika Heller, Tong Wang

Prior ratings, often used as quality signals in service management, can unintentionally introduce bias in follow-up evaluations. We conducted preregistered online experiments to test how prior ratings influence instructor evaluations when teaching quality is held constant. Our experimental evidence reveals sizeable negative effects of publishing low prior ratings, concentrated among inexperienced learners and on instructor-focused outcomes. Furthermore, male learners tend to correct unfairly low ratings only for male instructors, introducing a gender bias against female instructors. Learners exposed to high prior ratings tend to increase their ratings for instructors of the same gender, exhibiting confirmation bias. We also examine how different ratings displays influence evaluation bias and explore the underlying mechanisms. Our findings aim to identify operational strategies to mitigate bias, enhancing fairness and transparency in service delivery.

2 - Beyond Food Deserts: How Product Availability Explains Regional Disparities in Diet

Na Hyun Kim, The Wharton School, Philadelphia, PA, United States, Serguei Netessine, Dmitry Mitrofanov

Eating habits have long been correlated with socioeconomic status, fueling theories around food deserts—areas with limited access to nutritious foods—as key drivers of unhealthy diets. Such theories have significantly influenced federal policies and public health programs in the United States. However, recent research has challenged the assumption that eating habits are primarily supply-driven, suggesting consumer demand plays a more crucial role. This study revisits these claims by examining product availability. Product availability has been extensively studied in operations management and identified as pivotal to consumer behavior; however, its role in the context of healthy eating has received limited attention.

Leveraging a rich dataset from a leading U.S. online grocery delivery platform, we capture both stocking levels (product availability) and actual sales transactions. By assessing differences in healthy food access at the zip code level, our study explores heterogeneous impacts of product availability on grocery consumption and systematic differences in grocers' stocking behavior.

We hypothesize that current policies focusing on price subsidies, store openings, or healthier product assortments may fall short due to insufficient consideration of grocers' inventory stocking decisions. Ultimately, this research will provide novel insights on operational strategies of grocers operating in food-insecure/-secure markets and inform policymakers.

3 - Reputation and Reserve Dynamics of Non-Profit Organizations

Boyang Mu, CREST-Institut Polytechnique de Paris, Palaiseau, France, Mahyar Eftekhari

This study examines how nonprofit organizations (NPOs) can strategically manage financial reserves and reputation amid uncertainty. We develop a dynamic model in which key organizational decisions—such as spending on mission-driven programs and fundraising activities—affect both liquidity and perceived organizational credibility. These interdependent factors evolve dynamically, modulating donor contributions sensitive to both external conditions and internal actions. The model highlights trade-offs NPOs face: program expenditures amplify visibility and societal impact, yet excessive reserve accumulation or aggressive fundraising may erode donor confidence. The objective is to optimize long-term mission sustainability and operational continuity amidst financial volatility and reputational pressures. We also present empirical evidence to illustrate how these dynamics play out in practice. Our findings offer guidance for NPOs seeking to harmonize immediate impact with enduring resilience in resource-scarce settings. By optimizing reserve buffers and reputation management, NPOs can safeguard future operational capacity, fostering sustained donor trust and mission efficacy despite economic turbulence and evolving public sentiment.

4 - Process Configurations for Digital Government-to-Person (G2p) Payments and Uptake of Social Assistance

Rakesh Allu, Cornell University, Ithaca, New York, United States, Maya Ganesh, Sarang Deo, Sripad Devalkar

Government-to-Person (G2P) payments are being increasingly digitized in low- and middle-income countries. Traditionally, these payments were processed through a decentralized configuration, wherein each welfare program independently manages its funds and electronically credits payments to citizens. Lately, there has been a movement towards a centralized configuration, wherein funds across all welfare programs are pooled into a central bank account and disbursed using an account mapper that leverages citizens' universal identity. Despite the potential benefit of efficient fund utilization, such a centralized configuration can undermine citizen welfare owing to the techno-managerial complexity of its implementation due to the coordination required among multiple entities, the banks, the welfare programs, and the account mapper. We empirically examine the effect of configuration choice in the context of India's workfare program by assembling a unique dataset of 5.7 million digital payments made to 263,498 citizens over 30 months in one state, where 65% of the citizens gradually transitioned from a decentralized to a centralized configuration. Using this natural experiment, we find that centralization decreases average monthly work uptake per citizen by 50.3%. The key mechanism driving the decline is “redirection of wages”, i.e., disbursement of the wages to a different bank account under centralization compared to the one under decentralization. In our dataset, over 22% of citizens on centralized configuration experienced wage redirection and their work uptake was 11.36% lower compared to those without redirection. We further find that the reduction in work is larger among citizens whose wages were redirected to a geographically farther bank.

TD53

Building B Level 3 B313

Machine Learning Aided Designs in Operations

Invited Session

Revenue Management

Chair: Renfei Tan, Massachusetts Institute of Technology, Cambridge, MA, United States

Co-Chair: David Simchi-Levi, MIT, Cambridge, MA, United States

1 - Prediction-Guided Active Experiments

Hongyu Chen, Massachusetts Institute of Technology, Cambridge, MA, United States, David Simchi-Levi, Ruicheng Ao

In this work, we introduce a new framework for active experimentation, the Prediction-Guided Active Experiments (PGAE), which leverages predictions from an existing machine learning model to guide sampling and experimentation. Specifically, at each time step, an experimental unit is sampled according to a designated sampling distribution, and the actual outcome is observed based on an experimental probability. Otherwise, only a prediction for the outcome is available. We begin by analyzing the non-adaptive case, where full information on joint distribution of the predictor and the actual outcome is assumed. For this scenario, we derive an optimal experimentation strategy by minimizing the semi-parametric efficiency bound for the class of regular estimators. We then introduce an estimator that meets this efficiency bound, achieving asymptotic optimality. Next, we move to the adaptive case, where the predictor is continuously updated with newly sampled data. We show that the adaptive version of the estimator remains efficient and attains the same semi-parametric bound under certain regularity assumptions. Finally, we validate PGAE's performance through simulations and a semi-synthetic experiment using data from the US Census Bureau. The results underscore the PGAE framework's effectiveness and superiority compared to other existing methods.

2 - How to Sell a Service with Uncertain Outcomes

Alec Sun, University of Chicago, Chicago, IL, United States, Krishnamurthy Iyer, Haifeng Xu, You Zu

Motivated by the recent popularity of machine learning training services, we introduce a contract design problem in which a provider sells a service that results in an outcome of uncertain quality for the buyer. The seller has a set of actions that lead to different distributions over outcomes. We focus on a setting in which the seller has the ability to commit to an action and the buyer is free to accept or reject the outcome after seeing its realized quality. We propose a two-stage payment scheme where the seller designs a menu of contracts, each of which specifies an action, an upfront price and a vector of outcome-dependent usage prices. Upon selecting a contract, the buyer pays the upfront price, and after observing the realized outcome, the buyer either accepts and pays the corresponding usage price, or rejects and is exempt from further payment. We show that this two-stage payment structure is necessary to maximize profit: only upfront prices or only usage prices is insufficient. We then study the computational complexity of computing a profit-maximizing menu in our model. While computing the exact maximum seller profit is NP-hard even for two buyer types, we derive a fully-polynomial time approximation scheme (FPTAS) for the maximum profit for a constant number of buyer types. Finally, we prove that in the single-parameter setting in which buyers' valuations are parametrized by a single real number that seller revenue can be maximized using a menu consisting of a single contract.

3 - Proof-of-Learning with Incentive Security

Yuan Zhou, Tsinghua University, Beijing, China, People's Republic of, Zishuo Zhao, Zhixuan Fang, Xuechao Wang, Xi Chen, Hongxu Su, Haibo Xiao

Most concurrent blockchain systems rely heavily on the Proof-of-Work (PoW) or Proof-of-Stake (PoS) mechanisms for decentralized consensus and security assurance. However, the substantial energy expenditure stemming from computationally intensive yet meaningless tasks has raised considerable concerns surrounding traditional PoW approaches. The PoS mechanism, while free of energy consumption, is subject to security and economic issues. Addressing these issues, the paradigm of Proof-of-Useful-Work (PoUW) seeks to employ challenges of practical significance as PoW, thereby imbuing energy consumption with tangible value. While previous efforts in Proof of Learning (PoL) explored the utilization of deep learning model training SGD tasks as PoUW challenges, recent research has revealed its vulnerabilities to adversarial attacks and the theoretical hardness in crafting a byzantine-secure PoL mechanism. In this paper, we introduce the concept of incentive-security that incentivizes rational provers to behave honestly for their best interest, bypassing the existing hardness to design a PoL mechanism with computational efficiency, a provable incentive-security guarantee and controllable difficulty. Particularly, our work is secure against two attacks, and also improves the computational overhead. Furthermore, while most recent research assumes trusted problem providers and verifiers, our design also guarantees frontend incentive-security even when problem providers are untrusted, and verifier incentive-security that bypasses the Verifier's Dilemma. By incorporating ML training into blockchain consensus mechanisms with provable guarantees, our research not only proposes an eco-friendly solution to blockchain systems, but also provides a proposal for a completely decentralized computing power market in the new AI age.

4 - Experimental Design in the Presence of Supply Chain Spillovers

Yunhong Wang, National University of Singapore, Singapore, Singapore, Hanzhang Qin, Jinglong Zhao

In this paper, we develop a framework to design experiments for a regionalized supply chain, especially in the presence of supply chain spillover effect. Interferences are common in online platforms, and we identify a new interference pattern from the spillover effect of regionalization fulfillment, different from the previous setting in online platforms. We also propose different experimental designs and corresponding standard estimators for estimating the global treatment effect. We model customer purchase behavior via time-dependent choice model and use mean-field limit to capture the system dynamics. Finally, we provide sign and bias analysis for the standard inverse probability weighting estimator and propose some alternative estimators that can use more system information than the standard estimator.

5 - Imaged-Based Similarity for Demand Forecasting: a Novel Multimodal Method to Exploit Images&Rscquo; Latent Information

David Simchi-Levi, MIT, Cambridge, MA, United States, Junyi Sha, Hanwei Li, Yuxiang Liu, Michelle Wu

Demand forecasting for new fashion products is particularly challenging due to the lack of historical data. While companies often rely on textual information, few recent studies have explored incorporating product images as an additional data source. In this paper, we propose novel methods to utilize both textual descriptions and visual representations of fashion products, demonstrating their impact on forecasting new product sales. We examine different models to study the use of textual features, image features, and their combination for forecasting purposes. Specifically, we introduce a novel hybrid model that integrates an Autoencoder, a pre-trained Transformer, and the gradient boosting-based model Catboost. The data used in this study comes from a European fashion company. The model that shows the highest performance with purely textual information is selected as our baseline. Our proposed model leverages the concept of product similarity for image-based instance selection. This approach significantly improves prediction accuracy in certain fashion categories and enhances overall prediction accuracy by more than 8.5% in terms of both root-mean-square error (RMSE) and median absolute percentage error (MdAPE). A robustness test on another open fashion dataset, VISUELLE, confirms that our proposed models consistently improve demand forecasting accuracy in similar contexts. The hybrid model's incorporation of both textual and visual data significantly enhances the accuracy of demand forecasting for new fashion products. The innovative approach of using product similarity for image-based instance selection effectively addresses dataset imbalance issues, ensuring more reliable forecasts across various fashion categories.

TD54

Building B Level 2 B216

Advancements in Revenue Management and Discrete Choice Models: Theory and Applications

Invited Session

Revenue Management

Chair: Yunlong Wang, Shanghai Jiao Tong University, Shanghai

Co-Chair: Rui Zhang, University of Colorado Boulder, Boulder, CO, United States

1 - A Scalable Surrogate Approach for Multi-Purchase Assortment Optimization Under A General Random Utility Model

Wenhao Gu, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Anton Braverman, Tarek Abdallah

We consider the problem of multi-purchase assortment optimization under a general random utility model. We introduce a surrogate problem capable of handling random basket sizes and general utility distributions. The surrogate model approximates the original problem by leveraging asymptotic structure and remains effective even at moderate problem scales. We further develop algorithmic techniques—including dimension reduction via distributional clustering and a linear programming relaxation—to enable efficient optimization over large product sets and heterogeneous demand profiles. We provide rigorous performance guarantees for the surrogate and its relaxations, with provably vanishing error under mild conditions. Numerical experiments demonstrate strong empirical performance and significant runtime improvements over existing methods, particularly in large-scale settings.

2 - Assortment Optimization for Online Video Games

Yunlong Wang, Shanghai Jiao Tong University, Shanghai, China, People's Republic of

We consider an assortment optimization problem for a class of online video games where the in-game virtual store has a unique structure with two sections: Featured and Just For You (JFY). All customers (players) are offered the same Featured section assortment, whereas the JFY section is used for personalized recommendations. Our paper is the first to study assortment optimization for the gaming industry under discrete choice models; it is also the first to devise both exact and approximate solution approaches for the constrained mixture-of-nested-logit model. We model customer choice under a constrained mixture-of-nested-logit model and propose a number of solution methods for the resulting assortment optimization problems, including an exact mixed-integer nonlinear programming (MINLP) formulation and three approximate solution methods: a fully polynomial time approximation scheme (FPTAS), a mixed-integer linear programming (MILP) formulation, and a fast heuristic algorithm. We provide theoretical performance guarantees for the FPTAS, MILP formulation, and heuristic algorithm. Numerical experiments show that our approaches perform well across a variety of settings based on conversations with our industry partner to mimic the instances found in their video game stores. These experiments demonstrate that the MINLP can obtain optimal solutions efficiently. Our work provides guidance for how online video game stores can manage assortments for effective revenue maximization. We propose an "assortment" of solution approaches that allows practitioners to choose a method that best suits their environment. Furthermore, we conduct simulations based on ground-truth instances to demonstrate that considering the nests in the mixture-of-nested-logit model can better capture consumer behavior.

3 - Data Driven Decision Making from Censored Demand

Boxuan Zhou, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Lavanya Marla, Yunzong Xu

We address the single-period newsvendor problem with censored demand by developing a distributionally robust optimization framework that leverages empirical CDF bounds at sampled demand levels. Our ambiguity set imposes moment-type constraints on observed sales and censored observations to capture the complete data structure. Under a nonparametric assumption, we derive a closed-form robust order quantity via an integral equation that equalizes worst-case regret around the critical quantiles. By discretizing lower and upper CDF bounds at gridpoints, we obtain an explicit, efficiently computable implementation formula. The framework extends to parametric settings, enables quantile-convergence analysis, and paves the way for feature-based, multi-period inventory and pricing extensions under demand censoring.

4 - Play-to-Earn Blockchain Video Games: a Glimpse into the Future of Work

Lifei Sheng, University of Houston Clear Lake, Houston, TX, United States, Hubert Pun

“Play-to-Earn” games allow consumers monetize in-game achievements by trading leveled-up digital assets, thanks to the smart contract and traceability capabilities of blockchain. We model a video game publisher’s decision to allow tradability, analyzing its effects on consumer behavior and revenue. While making an asset tradable would introduce competition for the publisher, tradability may still be allowed even when it generates no trading revenue and incurs administrative costs. Our findings reveal strategic trade-offs for publishers and offer insights into broader applications of blockchain in digital markets.

TD55

Building B Level 3 B314

Innovation and New Frontiers in Retail Operations

Invited Session

Revenue Management

Chair: Alys Liang, McGill, Montréal, QC, Canada

Co-Chair: Alireza Alavi

1 - Assortment Optimization Without Prediction: An End-to-End Framework with Transaction Data

Shaoyu Wang, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Pin GAO, Andre Cire, Ningyuan Chen

Firms leverage assortment optimization to identify a set of products to offer to consumers in order to maximize the potential revenue. Traditionally, firms first use the historical transaction data to estimate a discrete choice model that predicts the choice behavior of future consumers, and subsequently solve the optimal assortment with such an estimated discrete choice model. This is typically referred to as the **predict-then-optimize** approach. In this paper, we propose an alternative approach that maps the data to the assortment decision in an **end-to-end** fashion, without the need to estimate a discrete choice model. The formulation extracts the partial orders induced from each customer in the historical transaction data. The assortment is then selected to maximize the worst-case revenue for a new customer when his/her preference is uniformly drawn from historical partial orders. We establish the NP-hardness of this problem and develop a mixed-integer programming formulation for exact solutions. Furthermore, we provide a number of approximation algorithms with theoretical guarantees.

To demonstrate the external validity of the framework, we show the recommended assortment can achieve 1/2 of the optimal revenue when the data is generated by a rank-based choice model.

The numerical result shows that the end-to-end framework is particularly attractive when (1) the data is scarce so that estimating model parameters can be unstable or (2) when the consumer behavior is complex and a pre-determined discrete choice model is likely to be misspecified.

2 - Joint Inventory Placement, Assortment Personalization, and Order Fulfillment for Substitutable Products

Mikhail Fadin, Cornell Tech, New York, NY, United States, Omar El Housni, Huseyin Topaloglu

Modern e-commerce platforms leverage distributed warehouse networks to store inventory and directly fulfill customer orders in near real time. A key challenge is coordinating (i) how to allocate inventories across network facilities subject to capacity and supply constraints; (ii) how to dynamically select personalized product assortments based on customer context and real-time stock levels; and (iii) how to route chosen orders to warehouses to maximize net profit. We formalize this joint optimization problem and give approximation algorithms. Extensive numerical experiments demonstrate that in practice our method achieves near-optimal net profits even for relatively low capacities and supply levels and, combined with simple additional heuristics, can perform even better and scale extremely well computationally.

3 - Service Design Using Personal Information

Bogyeom Lee, University of Michigan, Ross School of Business, Ann Arbor, MI, United States, Hyun-Soo Ahn, Joline Uichanco

We study how a firm should design personalized services when customers differ in both their valuation for base service and their sensitivity to add-ons. Due to regulatory and operational limits, firms often cannot observe both dimensions of heterogeneity. We model a monopolist choosing between collecting full information, partial information (on one dimension), or offering uniform service, and compare these to an indirect mechanism design approach. Our results show that partial information can often substitute for full information at a lower cost and can also outperform indirect mechanism design when information is verifiable. In addition, customers may voluntarily disclose private information when it improves service quality. When information is not verifiable, misreporting typically harms the firm but can also be beneficial by helping avoid market failure.

4 - Who Should Fulfill? the Role of Return Policies in E-Commerce Fulfillment Strategy

Serdar Simsek, The University of Texas at Dallas, Richardson, TX, United States, Ozalp Ozer, Jiarui (Gary) Wei

Third-party sellers on e-commerce platforms can fulfill their orders either through the platform’s own logistics service—Fulfillment by Platform (FBP)—or by managing fulfillment independently using carriers such as UPS or FedEx, referred to as Fulfillment by Seller (FBS). This paper examines how lenient return policies offered under FBP affect consumers’ purchase decisions and sellers’ revenues. We develop a stochastic model to capture the dynamics of consumer purchasing and return behavior under varying return policies. Lenient return policies can both increase product sales by reducing consumers’ perceived risk and raise return rates, potentially eroding seller profitability. We analyze how the relative value of FBP versus FBS varies with consumers’ return hassle cost, perceived value of FBP, product valuation uncertainty, and heterogeneity in consumer preferences. Our findings offer practical insights for sellers evaluating fulfillment options and for platforms seeking to design effective return policies and align their logistics services with suitable seller and product types.

TD56

Building B Level 3 B315

Auction and Platform Design

Invited Session

Auctions and Market Design

Chair: Yeon-Koo Che, Columbia University, New York, NY, United States

Co-Chair: Yeon-koo Che, Columbia University, New York, NY, United States

Co-Chair: Andrew Choi, University of Michigan, Ann Arbor, MI, United States

1 - Selling Heterogeneous Objects to Unit-Demand Bidders

Nina Bobkova, Rice University, Houston, TX, United States

We study sequential auctions of heterogeneous objects when bidders have unit demand. We characterize equilibrium bidding behavior when bidders observe their current valuation but face uncertainty about future objects. Bidders with low current values strategically delay participation to preserve the option value of competing for future

objects. Two opposing forces shape bidding behavior over time: the resource effect, as the value of waiting diminishes with fewer objects remaining, and the competition effect, as fewer active bidders increase expected payoffs in later rounds. We characterize the symmetric equilibrium in first-price, second-price, and English auctions, showing when revenue equivalence holds despite strategic delay. Compared to the classical homogeneous-object setting of Milgrom and Weber (1982), heterogeneous goods introduce nontrivial thresholds for participation. We discuss implications for the disclosure of information about future objects and the optimal order of sale.

2 - Optimal Platform Design

Cole Wittbrodt, Columbia University, New York, NY, United States

Search and matching increasingly takes place on online platforms. These platforms have elements of centralized and decentralized matching; platforms can alter the search process for its users, but are unable to eliminate search frictions entirely. I study a model where platforms can change the *distribution* of potential partners that an agent searches over and characterize search equilibria on platforms. When agents possess private information about their match characteristics and the platform designer acts as a profit maximizing monopolist, I characterize the optimal platform. If match characteristics are complementary and utility is transferable, I show that the solution to this screening problem is efficient, despite the presence of hidden information and market power. Matching under the optimal platform is perfectly assortative --- there is no equilibrium mismatch.

3 - Optimal Auction Design for Dynamic Stochastic Environments: Myerson Meets Naor

Andrew Choi, University of Michigan, Ann Arbor, MI, United States, Yeon-Koo Che

Allocation of goods and services often involves both stochastic supply and stochastic demand. Motivated by applications such as cloud computing, gig platforms, and blockchain auctions, we study the design of optimal selling mechanisms in an environment where buyers with private valuations arrive stochastically and are assigned goods that also arrive stochastically, and either buyers or goods can be held in a queue at costs until allocation. The optimal mechanism dynamically leverages competitive pressure across times by managing the queue of buyers and inventory of goods, using reserve prices that increase with queue length and an auction to allocate the goods.

4 - Optimal Auction Design for Dynamic Stochastic Environments: Myerson Meets Naor

Andrew Choi, University of Michigan, Ann Arbor, MI, United States, Yeon-Koo Che

Allocation of goods and services often involves both stochastic supply and stochastic demand. Motivated by applications such as cloud computing, gig platforms, and blockchain auctions, we study the design of optimal selling mechanisms in an environment where buyers with private valuations arrive stochastically and are assigned goods that also arrive stochastically, and either buyers or goods can be held in a queue at costs until allocation. The optimal mechanism dynamically leverages competitive pressure across times by managing the queue of buyers and inventory of goods, using reserve prices that increase with queue length and an auction to allocate the goods.

TD57

Building B Level 3 B316

Selected Auctions and Market Design Papers from the EC 2025 Conference II

Award Session

Auctions and Market Design

Chair: Rad Niazadeh, Chicago Booth School of Business, Chicago, IL, United States

Co-Chair: Christian Kroer, Columbia University, New York, NY, United States

Co-Chair: Jiaqi Lu, the Chinese University of Hong Kong (Shenzhen), Shenzhen, N/A

1 - Data-Driven Mechanism Design: Jointly Eliciting Preferences and Information

Marek Bojko, Yale University, New Haven, CT, United States, Dirk Bergemann, Paul Dütting, Renato Paes Leme

"We study mechanism design when agents have private preferences and private information about a common payoff-relevant state. We show that standard message-driven mechanisms cannot implement socially efficient allocations when agents have multidimensional types, even under favorable conditions. To overcome this limitation, we propose data-driven mechanisms that leverage additional post-allocation information, modeled as an estimator of the payoff-relevant state. Our data-driven mechanisms extend the classic Vickrey-Clarke-Groves class. We show that they achieve exact implementation in posterior equilibrium when the state is either fully revealed or the utility is affine in an unbiased estimator. We also show that they achieve approximate implementation with a consistent estimator, converging to exact implementation as the estimator converges, and present bounds on the convergence rate. We demonstrate applications to digital advertising auctions and large language model (LLM)-based mechanisms, where user engagement naturally reveals relevant information.

2 - Robust Dynamic Staffing with Predictions

Rad Niazadeh, Chicago Booth School of Business, Chicago, IL, United States, Yiding Feng, Vahideh Manshadi, Saba Neyshabouri

We consider a natural dynamic staffing problem in which a decision-maker sequentially hires staff over a finite time horizon to meet an unknown target demand at the end. The decision-maker also receives a sequence of predictions about the demand that become increasingly more accurate over time. Consequently, the decision-maker prefers to delay hiring decisions to avoid overstaffing. However, workers' availability decreases over time, resulting in a fundamental trade-off between securing staff early (thus risking overstaffing) versus hiring later based on more accurate predictions (but risking understaffing). This problem is primarily motivated by the staffing challenges that arise in last-mile delivery operations. A company such as Amazon has access to flexible gig economy workers (through Amazon Flex) whose availability decreases closer to the target operating day, but they can be hired at any time before that day if they are available. We study the above problem when predictions take the form of uncertainty intervals that encompass the true demand. The goal of the decision-maker is to minimize the staffing imbalance cost at the end of the horizon against any sequence of prediction intervals being chosen by an adversary. Our main result is the characterization of a simple and computationally efficient online algorithm that obtains the optimal worst-case imbalance cost; said differently, it is minimax optimal. At a high level, our algorithm relies on identifying a restricted adversary against which we can characterize the minimax optimal cost by solving a certain offline LP.

3 - A Stochastic Growth Model for Online Platforms

Farbod Ekbatani, Chicago Booth School of Business, Chicago, IL, United States, Rene Caldentey

We investigate the growth trajectory of online platforms operating as market makers in two-sided matching markets. Their expansion, often referred to as the "chicken-and-egg dilemma" in the literature, poses an inherent challenge; service providers find value in joining the platform due to the presence of customers, while customers are drawn to the platform by the availability of diverse services. This endogenous network effect influences the growth trajectory and propels emerging platforms towards "get big fast" type of strategies to rapidly achieve critical mass on both sides. Much of the existing literature tackles the chicken-and-egg dilemma using a static market-clearing approach, where network externalities determine a platform's resulting equilibrium size. However, this approach overlooks the inherent dynamics of gradually recruiting servers and attracting new customers. Embedded in this growth process are compensation schemes and the related intertemporal incentive-compatibility constraints that a platform must satisfy to sustain its expansion. These are fundamental aspects that static models cannot fully capture. To explicitly capture the stochastic dynamics of a platform's gradual growth, we model the system as a Markov process. Service providers join according to a Poisson process whose intensity is modulated by the platform's compensation structure. Meanwhile, customers are transitory, arriving according to a Poisson process with an intensity that is non-decreasing in the number of active servers, reflecting their preference for larger platforms. Under mild conditions, we characterize the optimal server compensation schemes the platform should adopt and establish upper bounds on how rapidly and extensively the platform can grow.

4 - Approximately Optimal Online Mechanism for Multiunit Demand Buyers with Post-Allocation Inspection

Xuanjie Li, Duke University, Durham, NC, United States, Alexandre Belloni, Ali Makhadmeh

We consider a mechanism design problem where an auctioneer allocates k units of an item to multiunit demand buyers arriving in an online fashion, each with a constant private marginal value. The auctioneer can inspect buyers allocated to learn their types and use this information to lower payments to reward truthful buyers. By using tools from the Calculus of Variations, we fully characterize the optimal mechanism for a single-buyer problem subject to an upper bound on the allocation expectation. In particular, we characterize the optimal allocation strategy as the solution of an ordinary differential equation and establish that it is a continuous and

monotonically increasing function of the buyer's types. With the solution to the single-buyer problem, and by using connections to the prophet inequality literature, we design an online mechanism that achieves $1 - \sqrt{\alpha_{\text{max}}}$ of the optimal (offline) revenue, where α_{max} denotes the highest fraction of all units a buyer requests.

5 - Battery Operations in Electricity Markets: Strategic Behavior and Distortion

Jerry Anunrojwong, University of Washington, Seattle, WA, United States, Santiago Balseiro, Omar Besbes, Bolun Xu

Electric power systems are undergoing a major transformation as they integrate intermittent renewable energy sources, and batteries to smooth out variations in renewable energy production. As privately-owned batteries grow from their role as marginal “price-takers” to significant players in the market, a natural question arises: How do batteries operate in electricity markets, and how does the strategic behavior of decentralized batteries distort decisions compared to centralized batteries? We propose an analytically tractable model that captures salient features of the highly complex electricity market. We derive in closed form the resulting battery behavior and generation cost in three operating regimes: (i) no battery, (ii) centralized battery, and (ii) decentralized profit-maximizing battery. We establish that a decentralized battery distorts its discharge decisions in three ways. First, there is quantity withholding, i.e., discharging less than centrally optimal. Second, there is a shift in participation from day-ahead to real-time, i.e., postponing some of its discharge from day-ahead to real-time. Third, there is reduction in real-time responsiveness, or discharging less in response to smoothing real-time demand than centrally optimal. We also quantify the impact of the battery market power on total system cost via the Price of Anarchy metric, and prove that it is always between 9/8 and 4/3. That is, incentive misalignment always exists, but it is bounded even in the worst case. We calibrate our model to real data from Los Angeles and Houston. Lastly, we show that competition is very effective at reducing distortions, but many market power mitigation mechanisms backfire, and

TD58

Building B Level 3 B317

Analytics Insights in Healthcare

Contributed Session

Contributed

Chair: Olzhas Zhorayev, George Mason University, Arlington, VA, United States

1 - Understanding What Keeps Indian Users Engaged with Femtech Apps: A Mixed-Methods Study

SHREYA MUKHERJEE, IIT Kharagpur, Kharagpur, India, SAINI DAS

The FemTech sector, which includes a range of technological innovations aimed at addressing women's health and wellness, has witnessed remarkable global growth in recent years. These technologies cover areas such as menstrual tracking, fertility monitoring, maternal health, and mental well-being. However, in the Indian context, where discussions around women's health and hygiene are often hindered by longstanding social taboos and cultural stigma, there is limited scholarly attention on what drives users to continue engaging with FemTech applications after initial adoption. This study seeks to fill that research gap by adopting a sequential two-stage mixed-methods approach. In the first phase, qualitative data were collected through in-depth interviews with Indian FemTech users, and the narratives were analyzed using inductive thematic analysis. The second phase involved quantitative testing of an integrated model, combining elements from the Technology Acceptance Model (TAM), Diffusion of Innovation (DOI) theory, and the emergent qualitative themes. Survey data collected from a diverse group of FemTech users across India was analysed to validate the model using SmartPLS. Findings from the study reveal that Perceived Usefulness (PU), Perceived Ease of Use (PEOU), and Perceived Enjoyment (PEN) play a critical role in shaping users' intentions to continue using FemTech apps. These results offer valuable implications for app developers, government agencies, and health policymakers, highlighting the need to prioritize user-centric design and culturally sensitive strategies to foster long-term engagement and improve women's health outcomes in India.

2 - Overcoming Vaccine Hesitancy by Multiplex Social Network Targeting

Marzena Fügenschuh, Berliner Hochschule für Technik, Berlin, Germany, Feng Fu

Understanding the impact of social factors on disease prevention and control is one of the key questions in behavioral epidemiology. The interactions of disease spreading and human health behavior such as vaccine uptake give rise to rich dueling dynamics of biological and social contagions. In light of this, it remains largely an open problem for optimal network targeting in order to harness the power of social contagion for behavior and attitude changes. Here we address this question explicitly in a multiplex network setting. Individuals are situated on two layers of networks. On the disease transmission network layer, they are exposed to infection risks. In the meantime, their opinions and vaccine uptake behavior are driven by the social discourse of their peer influence network layer. While the disease transmits through direct close contacts, vaccine views and uptake behaviors spread interpersonally within a long-range potentially virtual network. Our comprehensive simulation results demonstrate that network-based targeting with initial seeds of pro-vaccine supporters significantly influences the ultimate adoption rates of vaccination and thus the extent of the epidemic outbreak.

3 - Transformers with Koopman-Enhanced Graph Convolutional Network for Spatiotemporal Dynamics Forecasting

Zekai Wang, Fairfield University, Fairfield, CT, United States, Bing Yao

Spatiotemporal forecasting is essential for understanding and controlling dynamic systems in fields ranging from transportation to biomedicine. However, modeling these processes remains challenging due to irregular spatial geometries, nonlinear dynamics, and long-range temporal dependencies. In this study, we propose a novel two-stage predictive framework, Transformer with Koopman-enhanced Graph Convolutional Network (T-KGCN), to address these challenges. Our approach integrates graph convolutional networks (GCNs) to model complex spatial structures, Koopman operator theory to embed the system's state into a latent space with approximately linear dynamics, and a transformer to capture long-range temporal dependencies. Our contributions include: (1) a unified architecture combining GCN, Koopman embedding, and transformers; (2) a principled approach to model nonlinear dynamics with latent linear evolution; and (3) enhanced performance in long-range forecasting. While demonstrated in a cardiac setting, our framework is broadly applicable to diverse spatiotemporal prediction tasks.

4 - Uncovering Patterns of Comorbidity in Rheumatoid Arthritis: a Data Analytics Approach

Hamid Hadidi, Oklahoma State University, Stillwater, OK, United States, Kazi Arman Ahmed, Mohammad Fili, Dursun Delen, Guiping Hu, Lizhi Wang

Rheumatoid arthritis (RA) is a chronic autoimmune disorder affecting an estimated 1.5 million Americans and 18 million people worldwide, often accompanied by a complex array of comorbidities. Current understanding of the conditions that precede RA onset or develop subsequently remains limited. This investigation utilizes the extensive, longitudinal Oracle Real-World Data, formerly known as Cerner Real-World Data, to elucidate these patterns. We aim to: (1) identify potential antecedent conditions that may contribute to RA, and (2) characterize the spectrum of comorbidities emerging post-RA diagnosis. This study seeks to generate novel predictive insights which could be crucial for developing targeted interventions for high-risk individuals and for optimizing the management of RA patients by anticipating and mitigating comorbid conditions.

5 - AI-Assisted Systematic Review and Meta-Analysis in Diet-Microbiome-Kidney Research

Olzhas Zhorayev, George Mason University, Arlington, VA, United States, Meng-Hao Li, Xiaoyu Chen, Nithin Rogan Arulsamy, Naoru Koizumi

Chronic kidney disease (CKD) affects over 35 million U.S. adults and imposes a substantial burden on the healthcare system, as Medicare covers much of the costly treatments. Emerging research suggests that microbiome-targeted dietary interventions may slow disease progression and reduce healthcare burden. Existing evidence, however, remains fragmented and inconclusive. To address this gap, we introduce an AI-assisted tool that automates and accelerates a systematic review and meta-analysis of the diet-microbiome-kidney nexus by integrating web-scraping techniques and large language models (LLMs).

To manage the scope and complexity of the literature, we established a pipeline that automates each stage of the review process – literature search, screening, data extraction, and quality assessment. Our custom web-scraping tool retrieved 1,896 study records from Google Scholar, supplemented by 236 from PubMed (2000-2025). We then used OpenAI's LLMs to screen 2,132 records and identified 171 relevant studies, including 30 randomized controlled trials eligible for meta-analysis. Human

validation showed that GPT-o1 outperformed GPT-4o, achieving an 88% accuracy rate. Overall, LLM-based screening was 25 times more time-efficient than manual review. We further developed a local multimodal RAG-based tool to extract structured data on statistical measures, diet types, disease characteristics, and bacterial taxa for meta-analysis. Finally, study quality was assessed automatically, including evaluation of incomplete reporting, using predefined criteria.

This AI-assisted pipeline demonstrates substantial gains in scale, efficiency, and reproducibility of systematic reviews and meta-analyses. Beyond the CKD application, it is broadly adaptable and can support researchers, clinicians, and policymakers in generating timely, high-quality insights across diverse health domains.

TD59

Building B Level 3 B318

Optimization in Healthcare

Contributed Session

Contributed

Chair: Secil Sozuer, Lehigh University, Bethlehem, PA, 18015, United States

1 - Branch-Price-and-Cut for Patient Planning and Surgeon Scheduling

Babak Akbarzadeh, University of Gent, Gent, Belgium, Matthias Bogaert, Broos Maenhout

We address the problem of patient planning and surgeon scheduling in operating theatres, where patients must be scheduled alongside assigning time blocks to surgeons. The timing and length of the allotted time blocks depend on the patient characteristics on the surgeons' waiting lists. Solving this problem in an exact manner is challenging due to the large number of rooms, surgeons, and patients involved. To overcome this challenge, we propose an efficient branch-price-and-cut algorithm to find an optimal solution in an acceptable time span. For that purpose, we include different dedicated mechanisms to accelerate the solution-finding process. In this regard, the branch-price-and-cut tree is set up using an intelligent branching scheme, the nodes are searched in order of the lowest number of fractional variables, and improved bounds are computed to prune nodes earlier. To tighten the convex hull of the linear programming relaxation in each node, the algorithm relies on a row generation mechanism for adding valid inequalities. We conducted various computational experiments to demonstrate the performance of our algorithm and validate for each component the contribution of the implemented optimisation principles. Additionally, we show the superior performance of the proposed algorithm to alternative optimisation procedures.

2 - Regulation of Online Consultations Quality on Two-Sided Online Healthcare Platforms

Yumeng Zhang, Shandong University, Jinan, China, People's Republic of, Zizhuo Wang, Jianghua Zhang

As an emerging business model in the platform economy, online healthcare platforms have expanded rapidly. However, the information asymmetry between doctors and patients creates challenges for the quality of online consultations and threatens the stability of the online healthcare market. To prevent market failure, improving the regulation of online consultation quality while ensuring platform operations is a critical issue for online healthcare platforms. Through a game-theoretic model of a monopolistic online healthcare platform with two-sided market characteristics, we discuss the conditions for the platform's regulatory decision and the impact on price, the number of users, and patient surplus. Additionally, we described the equilibrium results under various scenarios. The analytical results show that increasing the level of platform regulation does not always lead to more patients or fewer doctors. This outcome is related to the losses patients experience from low-quality online consultations. Moreover, platform regulatory decisions depend on the losses to both patients and the platform from low-quality online consultations. When patient losses exceed a certain threshold, the platform becomes inoperable. Interestingly, our results also show that the optimal price, the number of users (patients and physicians), and patient surplus with regulation are not necessarily superior to those without regulation. This further confirms the significant impact of losses to patients and the platform from low-quality online consultations on platform operations. Moreover, depending on the extent of losses to both patients and the platform, either platform regulation or non-regulation can achieve a win-win situation for both the platform and patients.

3 - Appointment Scheduling Under Patient Preferences and Delay-Dependent No-Show

Hakan Kılıç, University of Toronto, Toronto, ON, Canada, Olga Bountali, Ningyuan Chen, Andre Cire

Patients with chronic diseases constitute a vulnerable group that has increased dramatically over the last decade. They require systematic medical attention and delay in their care can lead to severe health outcomes. These delays can arise from the limited capacity of the healthcare provider such that the patients cannot book an appointment on time or due to the patients not showing up to an appointment that was on time. To address these issues, we propose a Markov decision process framework that provides each patient a subset of appointments from which they can book, i.e., adjusted booking window, based on the patient's preferences regarding their appointment time. We leverage the booking window to manage both capacity and to mitigate no-shows, which are expected to increase with longer delays between appointment requests and appointment dates. We investigate the structure of the optimal policy and provide numerical insights.

4 - Pareto-Improving Bump Policies in Shared Operating Room Scheduling Systems

Hung Do, University of Vermont, Burlington, VT, United States, David Novak, Mitchell Tsai, Andrey Ukhov

For hospitals using shared operating rooms for scheduled and emergent cases, a bumping policy is needed to facilitate emergent cases when they arrive, but it often operates with competing objectives. The current bumping policies do not offer efficient access to capacity (i.e., Operating Rooms) as well as any mechanism for collaboration. We introduce, model and analyze a bumping policy called First-In-First-Out (FIFO) Bump Policy in the context of Operating Room Scheduling that facilitates more efficient use of capacity with pooling effect and offers a collaborative mechanism with potential positive behavioral effects to speedup service rates.

5 - Distributionally Robust Optimization for Appointment Scheduling in the Presence of Patient and Provider Lateness, Emergency, and No-Show Patients

Secil Sozuer, Lehigh University, Bethlehem, PA, United States, Karmel Shehadeh, Robert Storer

We address a single-provider scheduling problem given a fixed sequence of appointments with uncertain durations, patient unpunctuality, provider lateness, no-shows and emergencies. The aim is to optimize the decision of scheduling an arrival time for each patient, or equivalently, assigning appointment intervals between patients, along with first patient's arrival time. We remove the assumption that the probability distributions of the uncertain parameters in the scenario creation are known. To address this issue, we develop a distributionally robust optimization (DRO) model with an ambiguity set constructed to represent the uncertainty of the various distributions. Our DRO model finds the optimal appointment slots that minimize the worst-case expected operational cost over probability distributions residing in the ambiguity set. In our DRO model, we utilize a mean support ambiguity set that incorporates the mean, lower and upper bounds of the lateness, no-show, emergency, and appointment durations. We derive an equivalent mixed integer linear programming (MILP) reformulation of our proposed DRO model. We present the computational experiments along with the managerial insights.

TD60

Building B Level 3 B319

Machine Learning for Operations

Contributed Session

Contributed

Chair: Efe Sertkaya, University of Maryland, College Park, MD, United States

1 - Optimizing Multi-Objective Inventory Management Using Pareto-Conditioned Networks

SunJin Han, POSTECH, Pohang, Korea, Republic of, Dong Gu Choi, Daiki Min

Multi-objective inventory management problems have been extensively studied over the years. Given the trade-offs among multiple objectives, this study aims to approximate the Pareto front rather than seeking a single optimal solution. We employ a Pareto-conditioned network for the continuous review(r, Q) system. By leveraging the structural properties of the continuous review(r, Q) system, we propose a variation of the Pareto-conditioned network to improve the efficiency of identifying the Pareto front. The proposed method optimizes multiple objectives simultaneously rather than optimizing each objective individually. We compare the proposed algorithm against conventional evolutionary heuristic algorithms commonly used in prior research, demonstrating superior performance in terms of convergence speed and solution quality.

2 - A Reinforcement Learning Approach for Artillery Unit Routing and Target Scheduling with Resource Constraints

Jaewon Kim, Korea University, Seoul, Korea, Republic of, Hyungjoo Cha, DongKyun Kim, Euijin Lee, Taesu Cheong

This study presents a reinforcement learning (RL) framework for solving a complex routing and scheduling problem involving multiple artillery units in field operations. The problem involves coordinating multiple artillery units that must travel between nodes to engage enemy targets while minimizing the maximum mission completion time. Each unit begins from a depot, visits locations sequentially to engage different enemy targets, and returns to the depot upon mission completion. The system incorporates multiple ammunition types with varying ranges and limited inventory constraints. To address the computational complexity of this integrated routing and resource allocation challenge, we develop an online RL algorithm based on Masked Proximal Policy Optimization (Masked PPO) that masks invalid actions to ensure feasible decision-making in dynamic environments. Additionally, we implement curriculum learning by progressively increasing the difficulty of the simulation environment during training. The proposed RL approach is evaluated against traditional optimization models and Genetic Algorithm (GA) baselines across various scenarios. Experimental results demonstrate that the Masked PPO framework delivers robust solutions with competitive performance, particularly excelling in large-scale instances where optimization models exceed time limits and genetic algorithms fail to produce feasible solutions. These findings demonstrate the potential of reinforcement learning for complex military scheduling and routing applications requiring rapid adaptation and real-time decision-making.

3 - Procurement Lead TIME Estimation and Supplier Recommendation for Improving Supply Chain Resilience

Teppei Inoue, Hitachi America, Ltd., Santa Clara, CA, United States, Haiyan Wang

Recently, there have been a lot of supply chain disruptions caused by Covid 19, wars, and natural disasters, so supply chain resilience is getting attention. To improve the resilience, it is important to monitor and predict the supply chain operating parameters such as procurement Lead Time (LT) so that we can detect business environment change at an early stage. Furthermore, there is a risk that manufactures cannot receive parts and materials from their suppliers due to the disruptions, so it is necessary to be prepared with alternative suppliers. To address these issues, this presentation proposes procurement LT estimation using procurement transaction data and supplier recommendation using LLM and web scraping. In the LT estimation, we explore different modeling approaches to address many-to-many relationships among items, buyers, and suppliers, such as one model for any item-buyer-supplier tuple, one model for each item or multiple models for each item. Moreover, the models include external factors such as industrial production index and news articles to consider business environment changes. Based on the LT prediction, we find suppliers with supply delay risks and recommend alternative suppliers combining LLM and web search. The recommendation result includes the detailed information such as productionable products, ISO certification and sentiment analysis result for relevant news of each candidate to support optimal supplier selection. We demonstrate the effectiveness of our approach through the experiments with real procurement transaction data.

4 - Asymmetric Mapping with Wasserstein Loss: A Two-Stage Aae Framework for Classification Under Multi-Class Imbalance

JiSoo Hwang, Yonsei University, Seoul, Korea, Republic of, Chang Ouk Kim

In semiconductor manufacturing fault detection and classification (FDC) systems, defective samples are far fewer than normal ones, causing severe class imbalance and making the classification of defective and normal samples challenging. To ensure robust classification performance in such environments, oversampling methods are essential. In this study, we adopt an approach based on the Adversarial AutoEncoder (AAE) among existing oversampling techniques. We propose AMAAE-WL(Asymmetric Mapping Adversarial AutoEncoder with Wasserstein Loss). First, asymmetric mapping is applied to construct class-specific latent spaces for minority classes, thereby enhancing inter-class separability. Second, Wasserstein loss is incorporated to mitigate the training instability of AAEs, enabling more stable generative learning based on the distance between two distributions. Third, a step-wise decoder fine-tuning strategy using classification loss is introduced, ensuring that generated samples directly contribute to improving classification performance. The proposed method has the potential to generate high-quality, class-specific synthetic samples in multi-class imbalance scenarios while organically integrating the generative model with the classification model to maximize classification accuracy.

5 - Consistent Knn-Based Dataset Reduction with Optimal Weighting for Stochastic Optimization

Efe Sertkaya, University of Maryland, College Park, MD, United States, Alexander Estes

Dataset reduction techniques aim to construct a compact yet informative subset of a given dataset to improve computational efficiency while preserving statistical properties. Existing methods often rely on submodular optimization, leveraging heuristics such as gradient proximity or log-likelihood estimators to select representative data points. However, these approaches frequently lack formal consistency guarantees and do not provide optimal weighting schemes for use in downstream tasks, raising concerns about their reliability and efficacy. In this paper, we propose a novel dataset reduction method grounded in the k-nearest neighbors (kNN) framework and establish its asymptotic consistency. We show that, as the dataset size increases, the selected subset yields an estimator that converges to the true loss minimizer, thereby preserving the statistical fidelity of the full dataset. Crucially, our method also provides an optimal weighting scheme for the selected subset, which can be integrated into learning algorithms or optimization-based pipelines to improve generalization and solution quality. Extensive empirical evaluations on standard machine learning benchmarks and stochastic integer programming tasks demonstrate the practical performance of our approach.

TD61

Building B Level 4 B401

Multi-Modal Data Science for Precision Health

Invited Session

Quality, Statistics and Reliability

Chair: Meng Zhao, Lehigh University, Bethlehem, PA, United States

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

1 - Wasserstein Autoregressive Moving Average Model for Distribution-Valued Time Series Analytics

Yujing Huang, University at Buffalo, Buffalo, NY, United States, Xiaoyu Chen

Distribution-valued time series (DVTS), where each observation at time t is a probability distribution, are widely encountered. For example, in critical care settings such as intensive care units (ICUs), the daily distribution of systolic blood pressure (SBP) of each patient varies over time due to physiological changes, interventions, and patient heterogeneity. Most existing statistical tools are limited to data in vector space, hence being not suitable to analyze such DVTS. We propose a Wasserstein autoregressive moving average (WARMA) model based on Riemannian geometry to directly analyze DVTS. The merits of WARMA will be explored using ICU blood pressure data from the publicly available MIMIC-IV database.

2 - A Dual-Modal Large Language Model for Medical Time Series-Text Multimodal Learning

Ziyue Li, University of Cologne, Cologne, Germany

The recent rapid advancements in language models (LMs) have garnered attention in medical time series-text multimodal learning. However, existing contrastive learning-based and prompt-based LM approaches tend to be biased, often assigning a primary role to time series modality while treating text modality as secondary. We classify these approaches under a temporal-primary paradigm, which may overlook the unique and critical task-relevant information embedded in text modality like clinical reports, thus failing to fully leverage mutual benefits and complementarity of different modalities. To fill this gap, we propose a novel textual-temporal multimodal learning paradigm that enables either modality to serve as the primary while being enhanced by the other, thereby effectively capturing modality-specific information and fostering cross-modal interaction.

3 - Identifying Ptsd Through Multimodal Analysis of Sleep Dynamics

Cheng-Bang Chen, University of Miami, Coral Gables, FL, United States, Yujie Wang, Diane Lim

Accurate identification of Post-Traumatic Stress Disorder (PTSD) remains challenging due to subjective diagnostic criteria and overlapping symptoms with other conditions. Individuals cannot consciously alter sleep patterns, and sleep dynamics can reliably reflect health conditions, such as PTSD. This study hypothesizes that sleep, a complex dynamic process, can objectively indicate PTSD status. We developed a novel multimodal dynamic analysis framework utilizing polysomnography (PSG) signals to objectively detect PTSD. We leveraged state space representation to integrate Electrocardiogram (ECG), four-channel Electroencephalogram (EEG), and peripheral oxygen saturation (SpO₂) data to capture the complex, nonlinear, and nonstationary dynamics inherent in sleep. A multi-level heterogeneous recurrence quantification analysis (MHRQA) method was applied, analyzing sleep data from 108 participants. Hyperoctree Aggregate Segmentation was employed to enhance delineation within the state space, segmenting it into distinct regions. Sleep trajectories were transformed into symbolic sequences representing transitions across these segmented subspaces. Subsequently, an iterated function system projected these sequences into fractal circles, effectively capturing intricate sleep dynamics. Recurrence dynamics within each 30-second sleep epoch were quantified via MHRQA, generating rich dynamic features for identifying PTSD. An ensemble machine learning classifier classified epochs as PTSD-related or not. Performance validation using an 80/20 training-test split demonstrated high epoch-level accuracy (accuracy: 0.975, sensitivity: 0.964, specificity: 0.981, AUC: 0.996). Additional leave-one-out cross-validation indicated robust predictive performance at the patient level, achieving 0.75 accuracy. This multimodal dynamic analysis framework thus provides an objective, systematic, and reliable tool for PTSD detection based on sleep patterns, demonstrating considerable potential for clinical application and personalized care.

4 - Neurodegenerative Impact of Alzheimer's with Lewy Body Co-Pathology Characterized Through MRI-Based Brain Age Modeling with Deep Learning

Babak Ahmadi, University of Florida, Gainesville, FL, United States, Melissa Armstrong, Breton Asken, Mostafa Reisi, Abbas Babajani-Feremi

Alzheimer's disease (AD) and Lewy body (LB) pathology frequently co-occur, yet only recently have cerebrospinal fluid (CSF) α -synuclein seed amplification assays (SAA) enabled the in vivo detection of LB pathology, creating new avenues for examining its joint effects with AD on neurodegeneration. In this study, we developed a deep learning model trained on multi-cohort MRI data from 4,355 cognitively unimpaired individuals to estimate brain age, then applied it to 803 cognitively impaired participants. These individuals were categorized into four AD/LB pathology groups based on the p-tau181/A β 42 ratio for AD pathology and SAA status for LB pathology. Among these groups, the co-pathology subgroup (AD+LB+) demonstrated the greatest acceleration in brain aging, with saliency maps indicating more extensive neurodegeneration—a finding echoed by steeper longitudinal atrophy across multiple brain regions and associated cognitive declines. These findings underscore LB pathology's synergistic role in amplifying AD-related neurodegeneration, highlighting the importance of combined biomarker assays and targeted interventions for individuals harboring co-existing AD and LB pathology.

5 - Multimodal Learning for Health Data Integration with Missing Modality

Meng Zhao, Lehigh University, Bethlehem, PA, United States

Multimodal learning has emerged as a cornerstone of modern healthcare applications, integrating diverse data sources to build comprehensive patient representations for downstream clinical predictive tasks. However, in real-world clinical settings, one or more modalities are often absent due to cost constraints, sensor failures, or institutional data heterogeneity. In this work, we address this “missing modality” challenge by formulating it as a joint matrix recovery problem in which entire patient profiles (rows) may be unobserved in either data view. We develop a novel completion model which can jointly factorize all modalities and handle the row-wise data missing. Experiments demonstrate the benefits of our proposed method on multimodal data representation with incomplete modalities.

TD62

Building B Level 4 B402

AI/ML for Complex Engineering Systems

Invited Session

Quality, Statistics and Reliability

Chair: Zheng Liu, University of Michigan-Dearborn, 4901 Evergreen Rd, Urbana, 48128, United States

Co-Chair: Shancong Mou, University of Minnesota Twin Cities, Minneapolis, MN, 55455, United States

1 - Counterfactual Visual Explanation Process: Ranking and Utilizing Critical Features to Analyze a Machine Learning Model's Classification

Jian Hu, University of Michigan - Dearborn, Dearborn, MI, United States

Machine learning is increasingly utilized across diverse industries for automation, necessitating a clear understanding of model predictions to ensure accountability and foster user trust. In computer vision, counterfactual explanations provide insights by visually presenting changes in input images that significantly alter model outputs, thereby highlighting influential features. Our research introduces a novel approach to generate a series of visual explanations aimed at identifying critical, yet often-overlooked, features that impact model accuracy. We develop a ranking and feature reduction method to quantitatively assess the influence of these features on classification and prediction tasks. Our findings demonstrate that this approach not only enhances model transparency but also optimizes feature selection, thereby improving overall model efficiency. This work contributes to advancing the interpretability and reliability of machine learning models in practical applications.

2 - Utilizing Game Theory and Machine Learning in Policy Making

Gokce Dayanikli, University of Illinois at Urbana-Champaign, Champaign, IL, United States

In this talk, we introduce the Stackelberg mean field game problem to approximate the continuous-time, dynamic, and stochastic game between a principal and a large number of agents which can be used for different real life policy making applications such as epidemic mitigation. After introducing the general mathematical model, we discuss how to rewrite this bi-level problem as a single-level problem to propose an efficient numerical solution. In the model, the agents in the population play a non-cooperative game and choose their controls to optimize their individual objectives by interacting with the principal and other agents in the population through the population distribution. The principal can influence the resulting mean field game Nash equilibrium through incentives to optimize her own objective. After analyzing this game by using a probabilistic approach, we rewrite this intrinsically bi-level problem as a single-level problem and propose a machine learning approach to solve the Stackelberg mean field game. We conclude by discussing different applications such as mitigation of epidemics with non-pharmaceutical interventions and prevention of systemic risk in banking system with a regulator.

3 - Optimize Building Operations Using Data and AI

Gulai Shen, Harvard University, Cambridge, MA, United States, Ali Mehmani

Buildings are one of the largest energy consumers and carbon emitters. Due to the significant energy demand with increasing onsite generation and storage, buildings play a critical role in decarbonizing and regulating the electric grid. At the same time, the impact of indoor environment quality (IEQ) on occupants' wellness and productivity is substantial. These objectives together create complexity in finding optimal ways to manage building operations for reduced cost, energy, and emissions, together with better indoor environment and human-building interaction. Utilizing increased availability of data and AI technologies, we proposed novel frameworks, ways of

measuring, benchmarking, and controlling commercial office buildings with an occupant-centric approach. Case studies performed on buildings using the novel metrics like Indoor Environmental Wellness Index (IEW-Index) and PeopleHour show potential of improving building operations and enabling scalable AI solutions. In addition, AI algorithms with physics-infused hybrid models using predictive controls and reinforcement learning are implemented and maintained in real buildings. Results show 5% to over 10% energy and carbon reduction for test buildings, with over \$28M savings potential for the US General Service Administration (GSA) portfolio.

4 - Physics-Informed Machine Learning Enhanced Temperature Estimation of Pouch Cell with Indirect Liquid Cooling

Zheng Liu, University of Michigan-Dearborn, Dearborn, MI, United States

With the development of the indirect liquid cooling system for the pouch cell in energy storage systems, temperature estimation becomes an essential factor in design and operation. Traditional methods rely on experiments and finite element simulations, which are time-consuming and costly. Thus, data-driven methods are needed to accelerate the temperature estimation process while reducing the cost of designing and operating the energy storage systems. However, most data-driven methods require a large training dataset to achieve the desired accuracy, resulting in enormous computational costs for finite element simulation. Therefore, an approach to building a high-fidelity surrogate model with less training data is needed. This research proposes a physics-informed machine learning enhanced surrogate model to estimate the temperature of the pouch with the indirect liquid cooling. By integrating the information of the heat transfer partial differential equation into the loss function, the proposed surrogate model can achieve a high-fidelity temperature estimation with less training data. As a result, the proposed method achieved high-fidelity temperature estimation with over 50% reduction of the training data.

5 - Dynamic Parameter Calibration of Digital Twins Using Surrogate-Based Kalman Filtering

Jaeshin Park, University of Michigan, Ann Arbor, MI, United States, Chenfei Li, Eunshin Byon, William Bailey, Quazi Hussain, Yashodeep Lonari

Digital twins require accurate parameter estimation to remain consistent with their physical systems. We propose a dynamic calibration method that integrates a Kalman filter with a surrogate model, enabling real-time updates without costly simulator runs. The approach is demonstrated through a digital twin for thermal climate control in electric vehicles.

TD63

Building B Level 4 B403

AI and Machine Learning

Contributed Session

Contributed

Chair: Rang Gong, The Ohio State University, 2100 Neil Ave, Columbus, OH, 43210-1132, United States

1 - Classifying and Contextualizing the Unknown: a Continuous Abatement-Based Algorithm for Open Set Recognition

Steven Wang, University at Buffalo, Buffalo, NY, United States, Kyle Hunt, Zachary Steever, Chase Murray, Junsong Yuan, Mark Karwan

Open set recognition (OSR) tasks require classifiers to both assign labels to known classes and identify inputs from unknown categories. Traditional OSR approaches often rely on thresholding mechanisms and fail to provide insight into how unknown samples relate to known data. This paper introduces a novel threshold-free OSR framework that integrates a convolutional neural network with a continuous abatement model. The abatement model modifies class probabilities based on nearest-neighbor similarity in learned feature space, producing $|K| + 1$ confidence values—one for each known class and one for the open set. This formulation allows unknown inputs to be interpreted as distributions over both known and unknown regions of feature space, facilitating both detection and contextualization. We evaluate the framework using a synthesized benchmark constructed from MNIST and EMNIST and show that it outperforms a leading threshold-based OSR method across multiple performance metrics. Beyond image data, we extend our approach to contextualize unknown instances in the domain of integer programming classification, demonstrating the flexibility and generalizability of the method. By removing the need for arbitrary thresholds and enabling interpretability for unknowns, this work contributes a new perspective to the OSR literature and offers a practical tool for open-world machine learning applications.

2 - Mapping Industry-Wide Concerns in Hemp Agriculture by Applying Nlp to Public Sentiment on Youtube

Yihan Huang, North Carolina State University, Raleigh, NC, United States, Md Abdul Quddus, David Suchoff, Md Fashiar Rahman

Industrial hemp is now emerging as a valuable crop with wide applications in textiles, construction, and sustainable agriculture. However, serious obstacles still exist in the industry, including strict regulatory environments, inconsistent policies, market volatility, and limitations in processing infrastructure. This study how different hemp stakeholders express their experiences and concerns through public online platforms. By using natural language processing methods, we analyze over twenty-one thousand comments posted on YouTube videos related to industrial hemp. The analysis reveals two dominant themes. First, there is clear frustration surrounding legal restrictions, particularly the low THC threshold and conflicting regulations across regions. Second, many comments reflect growing optimism about the future of hemp, especially in sustainable building materials, technical textiles, and low-input farming systems. To better understand these discussions, we applied advanced sentiment analysis and topic modeling tools to extract emotion, intent, and major talking points from the comments. We also developed a statistical model showing a significant relationship between public sentiment and actual hemp production data, suggesting that online discourse can serve as an early indicator of industry performance. This research highlights the value of combining digital communication and data science to monitor agricultural sectors in real time. The findings offer practical insights for farmers, investors, and policymakers seeking to understand and respond to the rapidly developing hemp market.

3 - Adaptive Imputation Sensor Data with Scheduled Shutdowns

Mohammad Maydanchi, Auburn University, Auburn, AL, United States, Gregory Purdy, Daniel Silva

This study proposes an agentic imputation framework for time series sensor data that selects the most effective imputation model by evaluating its performance. The framework compares three candidate models: SAITS, a recurrent neural network (RNN)-based imputer, and a vector retrieval-driven imputer inspired by Retrieval-Augmented Generation (RAG). To determine the most suitable model for each missing value, the system retrieves similar historical patterns, masks the corresponding data points, and computes standard error metrics such as MAE, MAPE, and RMSE. The model that consistently achieves the lowest aggregated error across the retrieved patterns is then applied to impute the original missing value. Beyond this adaptive selection process, the study also conducts a direct comparative analysis of the three models under varying temporal and structural conditions. Experimental results show that the proposed framework improves imputation accuracy and demonstrates robustness across diverse time series behaviors.

4 - The Trap of Complexity in Experimentation

Yudi Huang, Kellogg School of Management, Northwestern University, Evanston, IL, United States, Sebastien Martin, Zhiwei Qin

In the tech industry, firms increasingly rely on continuous experimentation to guide system updates and improve performance. This paper reveals a subtle, often overlooked cost of such improvements: certain changes that yield immediate gains can increase the system's complexity, ultimately slowing future innovation. Complexity imposes two distinct burdens. First, it diminishes the likelihood of discovering additional successful changes; second, it forces the adoption of more intricate and time-consuming experimental designs, reducing the total number of experiments that can be conducted. Crucially, these future slowdowns remain invisible to standard experimental evaluations, which focus solely on immediate treatment effects.

By modeling this dynamic as a Markov Decision Process, we uncover a surprising pattern. When complexity is low, the firm naturally guards against increasing it. Yet once complexity surpasses a certain threshold, the firm's best response—despite full awareness of the long-term consequences—is to continue implementing complexity-increasing changes, locking itself into a high-complexity, low-innovation regime. Early-stage companies, driven by near-term goals, are especially susceptible to complexity "debt" that undermines their long-run adaptability. These findings caution that while quickly adopting profitable updates may seem optimal in the short term, preserving the ability to innovate at scale demands proactively managing complexity. In doing so, managers can balance immediate opportunities against the need to sustain experimentation and growth over the long term.

5 - Federated Wasserstein Barycenter by Dual Decomposition

Zhengqi Lin, Rutgers University, Piscataway, NJ, United States

We propose an efficient federated dual decomposition algorithm for calculating the Wasserstein barycenter of several distributions. The algorithm does not have access to local data and uses only highly aggregate information. The algorithm exhibits very low complexity of each iteration and significant scalability. We illustrate its virtues on several examples of mixture models.

6 - When More Isn't Better: Understanding the Marginal Impact of Robot Deployment on Campus Food Delivery Performance

Rang Gong, The Ohio State University, Columbus, OH, United States, Xiang Wan

Autonomous delivery robots are increasingly used to enhance last-mile logistics by reducing labor costs and increasing flexibility. However, it remains unclear whether simply adding more robots always leads to better operational performance. This study examines the impact of increasing the number of autonomous robots on delivery efficiency within a campus food delivery system. We utilize a comprehensive dataset that includes delivery logs from a robot company, class and event schedules, and weather data. To estimate the causal relationship between the number of robots and delivery time, we apply a double machine learning approach. In addition to estimating the overall impact, we investigate how this relationship may vary depending on the number of robots already existing in the system and the level of road congestion. Although the study is grounded in a university campus setting, the methodology and insights aim to inform broader applications of autonomous delivery systems, such as urban parcel logistics and drone-based delivery. This research contributes to the literature on automation and operational management by challenging assumptions of straightforward scalability and by offering guidance for more effective deployment and infrastructure planning.

TD64

Building B Level 4 B404

Environmental and Socially Responsible Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Sytse Wijnsma, UC Berkeley Haas, Berkeley, CA, United States

1 - Design of Points-Based Systems

Elisabeth Paulson, Harvard Business School, Somerville, MA, United States

This work considering the design of points-based systems commonly used in public sector contexts to allocate scarce resources such as public housing, organs, and visas. Points-based systems rank applicants according to a score, which is the weighted sum of a combination of indicator variables. The weight on a certain feature can be interpreted as the "points" assigned to that feature. In this paper, we provide a framework for determining an optimal score function, taking into account both the accuracy of the function and the cost of including each feature. In terms of feature costs, we consider the cost to applicants of acquiring/reporting the feature, as well as the cost to the administrator of verifying each feature.

2 - Labor Cost, Skills, and Productivity in Circular Operations

Cerag Pince, Quinlan School of Business, Loyola University Chicago, Chicago, IL, United States

As organizations pursue circular initiatives, they face non-traditional challenges in scaling operations. Among the most pressing concerns is labor cost, given that product recovery processes often require manual labor of varying skill levels. However, the variability in input flows and the difficulty of standardizing processes lead organizations to make complex trade-offs in selecting workforce capacity. Motivated by our interactions with a re-X service provider, we examine how circular organizations should determine their labor capacity. The unique features of the re-X context allow us to uncover novel insights into the interplay between labor costs, skill composition, productivity, and the role of process standardization. Our findings address a gap in the sustainable operations literature and offer actionable guidance for managing the distinct challenges of circular operations.

3 - Reducing Food Waste: the Impact of Online Grocery Shopping

Jimin Park, Massachusetts Institute of Technology, Cambridge, MA, United States, Yanchong Zheng

With online grocery shopping growing rapidly in recent years, an important question arises as to how it may impact food waste at the consumer and retail levels. This paper combines an analytical model and numerical analysis calibrated with real-world data to answer this question. In particular, we develop a bi-level optimization model in which the lower level captures consumers' repeat purchase decisions, and the upper level captures the retailer's inventory management problem. A key methodological contribution of our work is to endogenize consumers' repeat purchase decisions and translate them into the stochastic demand faced by the retailer. To tackle this complexity, we establish the unimodality of the retailer's inventory optimization problem within a perishability threshold that is satisfied by most products in practice. Building on this insight, we are able to characterize an asymptotically optimal base-stock policy for the retailer's problem. Our results show that online grocery shopping has the potential to reduce food waste at both the consumer and retail levels, but this potential critically depends on targeting the "right" consumers and products. To counter the increased risk of receiving less fresh products, online retail must enable more frequent purchase by lowering access costs and prioritize products with lower perishability and consumption. Under this strategy, the retailer can also decrease the safety stock held in the system, thereby reducing food waste also at the retail level. These insights highlight that online shopping can create an alignment between the reduction of food waste and greater cost efficiency for the retailer.

4 - Payments for Ecosystem Services: Balancing Upfront and Ex-Post Payments to Overcome Financial Barriers

Zhuoyang Liu, Stanford University, Stanford, CA, United States, Itai Ashlagi, Dan Iancu

Payments for Ecosystem Services (PES)—also known as payments for environmental services—are voluntary transactions in which providers receive compensation conditional on complying with prescribed natural resource management practices. PES schemes have proliferated as tools for environmental conservation and rural economic development in low-income regions with weak regulatory enforcement. Despite their popularity, PES programs suffer from severe adverse selection, rewarding landholders who would have conserved regardless of incentive payments. Moreover, by disbursing funds only after participants have successfully completed costly conservation actions, traditional PES excludes the most financially constrained providers, limiting overall effectiveness and skewing participation toward wealthier landholders.

Although there is ample evidence of adverse selection and financial barriers in PES implementation, no prior work systematically explores how to structure PES payments that mitigate both issues simultaneously. We fill this gap by studying how to optimally design two-stage PES schemes that incorporate both upfront and ex-post payments. Specifically, we develop a novel *two-stage* principal-agent screening model, grounded in contract theory, to analyze the interaction between a resource-constrained ecosystem service provider (the agent) and a buyer (the principal). We characterize the optimal contract structures that balance liquidity support and incentive compatibility, derive managerial insights for improving PES scheme design through coordinated upfront and performance-contingent payments, and calibrate our model with field data to demonstrate the validity and effectiveness of these insights.

5 - Interpretable Data-Driven Policy for Index-Based Yield Protection

Jingsheng Song, Duke University, Durham, NC, United States

Index-based yield protection is an alternative to traditional crop insurance that compensates farmers based on a predefined index (such as rainfall) rather than on actual yield losses. In this research, we propose a data-driven framework to determine the optimal subsidy payment under a government index-based yield protection policy. Our approach requires only minimal changes to widely adopted policies but significantly improves effectiveness. Specifically, it addresses two key limitations of existing policies: First, it accounts for the distinct impact of yield prediction errors on risk-averse farmers across different yield levels, enhancing the robustness to model misspecification. Second, it explicitly incorporates index accuracy when determining payments to farmers. In addition, we provide high-probability convergence guarantees for the parameter estimates used in our algorithm, and extend these to performance guarantees in terms of net benefit (i.e., total farmer surplus minus government expenditure). Finally, we calibrate our model using real-world data to demonstrate the advantages of our proposed approach.

TD65

Building B Level 4 B405

Foundations of AI

Invited Session

Artificial Intelligence

Chair: Siva Theja Maguluri, ISyE Georgia Tech, Atlanta, GA, United States

Co-Chair: Matthew Faw, Georgia Institute of Technology, Atlanta, GA, United States

1 - Gold-Medalist Performance in Solving Olympiad Geometry with Alphageometry2**Hoang Nguyen, Georgia Institute of Technology, Atlanta, GA, United States**

We present AlphaGeometry2, a significantly improved version of AlphaGeometry introduced in (Trinh et al., 2024), which has now surpassed an average gold medalist in solving Olympiad geometry problems. To achieve this, we first extend the original AlphaGeometry language to tackle harder problems involving movements of objects, and problems containing linear equations of angles, ratios, and distances. This, together with support for non-constructive problems, has markedly improved the coverage rate of the AlphaGeometry language on International Math Olympiads (IMO) 2000–2024 geometry problems from 66% to 88%. The search process of AlphaGeometry2 has also been greatly improved through the use of Gemini architecture for better language modeling, and a novel knowledge-sharing mechanism that enables effective communication between search trees. Together with further enhancements to the symbolic engine and synthetic data generation, we have significantly boosted the overall solving rate of AlphaGeometry2 to 84% for all geometry problems over the last 25 years, compared to 54% previously. AlphaGeometry2 was also part of the system that achieved silver-medal standard at IMO 2024 <https://dpmd.ai/imo-silver>. Last but not least, we report progress towards using AlphaGeometry2 as a part of a fully automated system that reliably solves geometry problems directly from natural language input.

2 - Tight Finite TIME Bounds of Two-Time-Scale Linear Stochastic Approximation with Markovian Noise**Shaan Ul Haque, Georgia Institute of Technology, ATLANTA, GA, United States, Sajad Khodadadian, Siva Theja Maguluri**

Stochastic approximation (SA) is an iterative algorithm for finding the fixed point of an operator using noisy samples and widely used in optimization and Reinforcement Learning (RL). The noise in RL exhibits a Markovian structure, and in some cases, such as gradient temporal difference (GTD) methods, SA is employed in a two-time-scale framework. This combination introduces significant theoretical challenges for analysis.

We derive an upper bound on the error for the iterations of linear two-time-scale SA with Markovian noise. We demonstrate that the mean squared error decreases as $\text{Trace}(\Sigma^k)/k + o(1/k)$ where k is the number of iterates, and Σ is an appropriately defined covariance matrix. A key feature of our bounds is that the leading term, Σ , exactly matches with the covariance in the Central Limit Theorem (CLT) for the two-time-scale SA, and we call them tight finite-time bounds. We illustrate their use in RL by establishing sample complexity for off-policy algorithms, TDC, GTD, and GTD2.

A special case of linear two-time-scale SA that is extensively studied is linear SA with Polyak-Ruppert averaging. We present tight finite time bounds corresponding to the covariance matrix of the CLT. Such bounds can be used to study TD-learning with Polyak-Ruppert averaging.

3 - A General-Purpose Theorem for High-Probability Bounds of Stochastic Approximation with Polyak Averaging**Sajad Khodadadian, Virginia Tech, Blacksburg, VA, United States**

Polyak–Ruppert averaging is a widely used technique to achieve the optimal asymptotic variance of stochastic approximation (SA) algorithms, yet its high-probability performance guarantees remain underexplored in general settings. In this paper, we present a general framework for establishing non-asymptotic concentration bounds for the error of averaged SA iterates. Our approach assumes access to individual concentration bounds for the unaveraged iterates and yields a sharp bound on the averaged iterates. We also construct an example, showing the tightness of our result up to constant multiplicative factors. As direct applications, we derive tight concentration bounds for contractive SA algorithms and for algorithms such as temporal difference learning and Q-learning with averaging, obtaining new bounds in settings where traditional analysis is challenging.

4 - Online Learning with Limited Switches (and Private Online Learning)**Karan Singh, Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, United States**

Online algorithms succeed by adapting to new information, but changing decisions often incur significant deployment costs. In this talk, we will see that folklore solutions to this problem are extraordinarily brittle.

Instead, we will address this problem in the online convex optimization framework, where learners can revise their decisions only a few times. We will give near-optimal algorithms for this setting. The resulting regret guarantee captures the *dimension-free* nature of regret bounds typical to OCO and improves prior results.

This problem connects directly to differentially private online learning, where each decision update risks privacy loss. As a key application, we will see the first private online learning algorithm that matches optimal *non-private* regret bounds for sub-constant privacy regimes.

TD66

Building B Level 4 B406

Causal Inference and Predictive Modeling in Complex Systems

Contributed Session

Contributed

Chair: Jungwoo Kim, KAIST, Yuseung-gu, Korea, Republic of

1 - Data-Secure Dual Transfer Learning from Heterogeneous Low-Rank and Sparse Panel Var Model**Yuchen Xu, University of California, Los Angeles, Los Angeles, CA, United States, George Michailidis**

Transfer learning (TL) enhances high-dimensional inference by borrowing strength from auxiliary data sets. We extend TL to multivariate time series settings through a panel Vector Autoregression (VAR) model, whose coefficients decompose into heterogeneous low-rank and sparse components. The proposed algorithm treats the two components independently and automatically detects the transferable auxiliary sources for each. Importantly, it operates solely on pre-trained estimators from the auxiliary panel, thereby preserving data privacy. We establish that the source detection mechanism achieves high accuracy under mild technical conditions. The proposed TL approach delivers significant performance improvements in data-scarce regimes—precisely the scenarios that motivate transfer learning. Extensive numerical experiments based on synthetic data, conducted for both correctly specified and misspecified time series models, along with applications to macroeconomic time series, confirm the practical effectiveness of our method. The results highlight the potential of TL to enhance predictive tasks in complex multivariate time series models when primary data are limited.

2 - Weighted Data Valuation for Statistical Learning via Nonconvex Programming Methods**Hassan Nojavan, Southern Methodist University, Dallas, TX, United States, Miju Ahn, Chul Moon**

We propose a new formulation of empirical risk minimization that accounts for the weights of data points. The proposed formulation aims to simultaneously estimate the model parameters while selecting significant data points. We propose an alternate weakly-convex search algorithm, which is a generalization of the alternate convex search method, and show the convergence results. The proposed method is applied to regression for outlier detection and the top N recommender system problems for data

valuation. Our numerical experiments demonstrate that the proposed approaches deliver high-quality solutions, outperforming some of the existing methods used in practice, while effectively identifying data anomalies.

3 - Tracing the Impact of Macro-Shocks on Annual S and P 500 Direction with Tree-Augmented Naïve Bayes

Taha Gulal, Pioneer Academy, Wayne, NJ, United States, Pervin Erdem, Arjon Bushaj, Serhat Simsek

As events like wartime mobilizations, global pandemics, presidential transitions, and recessions unfold, their immediate ripple effects across key macroeconomic indicators remain difficult to disentangle. Conventional analyses often consider these shocks piecemeal, missing how each one directly reconfigures factors such as interest rates, inflation, unemployment, and investor sentiment. In this study, we deploy a data-driven Tree Augmented Naïve (TAN) Bayes methodology to pinpoint which variables each high-impact event most strongly and directly influences—and to clarify how those links, in turn, steer the S&P 500's annual direction (“Up” vs. “Down”). The resulting model is parsimonious yet preserves accuracy, sensitivity, and specificity, providing investors, risk managers, and policymakers with a transparent map of event-factor pathways that can guide year-over-year asset allocation and contingency planning.

4 - Digital Transformation Disclosures and Inventory Volatility: Evidence from Text Analysis of 10-K Filings

Can Akbas, Southern Illinois University Edwardsville, Edwardsville, IL, United States, Stanislaus Solomon, Minseok Kim

This study investigates whether firms that emphasize digital transformation (DT) in their annual reports exhibit greater supply chain stability, as measured by inventory volatility. We develop a novel firm-year DT intensity score by applying a curated TF-IDF-based dictionary to 10-K filings for U.S. public companies from 2011 to 2023. This textual measure is then matched with Compustat quarterly inventory data to construct intra-year inventory volatility, a key operational indicator of stability and resilience.

Our empirical strategy employs panel regression models with firm and year fixed effects, incorporating standard controls (firm size, leverage, ROA) and supply chain–relevant factors (inventory intensity, demand uncertainty). The study also examines how the DT–volatility relationship varies across uncertainty levels, inventory intensity, and industry type. This research contributes to the growing intersection of supply chain management, digital transformation, and textual analytics. By grounding DT measurement in firm disclosures and linking it to operational outcomes, the study offers a scalable approach to capturing digital strategy execution.

5 - Data Fusion for Predicting Operational Speed

Julio Jurado, University of Oklahoma, Norman, OK, United States, Ali Shirazi

Operational speed plays a critical role in various traffic management activities, such as capacity estimation, safety improvement, and travel time reliability. Traditional models typically rely on statistical approaches and limited data sources. However, the growing availability of high-resolution, multi-source data allows for the development of more accurate predictive models. Leveraging this data requires advanced models capable of integrating and analyzing diverse inputs. This study presents a framework that aggregates data from various sources and uses deep neural networks to estimate traffic speed. Two datasets collected on Interstate 80 (I-80) in Utah and Wyoming are used for modeling. The data includes high-resolution data from speed sensors (aggregated at 5-minute intervals for Utah and 2-minute intervals for Wyoming), weather stations (10-minute intervals), and roadway features. Separate neural network models are trained for each dataset to capture region-specific patterns. These models are then integrated into a stacked model to improve the prediction power using a unified dataset. Evaluation results show that the proposed framework reasonably predicts operational speed under varying conditions.

6 - Combinatorial Latent Class Choice Model for Capturing Multidimensional Preference Heterogeneity

Jungwoo Kim, KAIST, Daejeon, Korea, Republic of, Taesik Lee

Latent Class Choice Models (LCCMs) are widely used to capture taste heterogeneity in discrete choice analysis. However, their single-membership structure constrains all utility coefficients to share a common latent segmentation, leading to three major issues: *oversplitting* (duplicated coefficients for near-identical effects), *undersplitting* (compression of heterogeneity into too few classes), and *misalignment* (inability to accommodate variable-specific segmentations). To relax these restrictions, we propose the *Combinatorial Latent Class Choice Model* (CLCCM). CLCCM first decomposes the utility function into interpretable components and assigns each component an independent latent class structure, so that the overall utility is constructed by combining the class memberships chosen for each component. We develop an EM-based estimation algorithm and apply the model to the data from a discrete choice experiment on breast cancer screening policies in Belarus. Compared with standard LCCMs, CLCCM delivers superior goodness-of-fit, more precise coefficient estimates, offering enhanced insights into class membership and estimated coefficients. These advantages highlight CLCCM's practical utility in uncovering nuanced preference structures, enabling more targeted, data-driven interventions.

TD67

Building B Level 4 B407

Statistical Foundations and Insights for Advancing Artificial Intelligence

Invited Session

Data Mining

Chair: Yiling Xie, Georgia Tech, Atlanta, GA, United States

Co-Chair: Tian-Yi Zhou, Georgia Institute of Technology, Atlanta, GA, 30318, United States

1 - Quantifying Discrimination in Outcome Tests

Hans Gaebler, Harvard University, Cambridge, MA, United States

The robust outcome test (Gaebler & Goel, 2025) is a simple and statistically sound strategy for measuring discrimination: If lenders approve fewer loans to one group, but that group repays its loans more often, it suggests lenders subjected that group to a stricter standard, only lending to its most qualified members. This simple logic can reveal that lenders have a double standard, but not how different the standards are. Researchers have developed alternative methods that can quantify the degree of discrimination, but with important tradeoffs: opaque statistical assumptions, restrictive data requirements, and often implausible expectations about decision maker rationality. In this talk, I'll introduce a modified robust outcome test that quantifies discrimination by transparently incorporating prior beliefs about how predictable the outcome (e.g., loan repayment) is. Using our method, we analyze 68 million traffic stops from across the U.S., finding evidence of widespread discrimination—including many instances the robust outcome test would have missed. Across 34 U.S. cities and states, we estimate that 10% to 25% fewer searches of minority drivers would occur under uniform evidentiary standards.

2 - Dynamic Contextual Pricing with Doubly Non-Parametric Random Utility Models

Lan Gao, University of Tennessee Knoxville, Knoxville, TN, United States

In the evolving landscape of digital commerce, adaptive dynamic pricing strategies are essential for gaining a competitive edge. This paper introduces novel doubly nonparametric random utility models that eschew traditional parametric assumptions used in estimating consumer demand's mean utility function and noise distribution. Existing nonparametric methods like multi-scale Distributional Nearest Neighbors (DNN and TDNN), initially designed for offline regression, face challenges in dynamic online pricing due to design limitations, such as the indirect observability of utility-related variables and the absence of uniform convergence guarantees. We address these challenges with innovative population equations that facilitate nonparametric estimation within decision-making frameworks and establish new analytical results on the uniform convergence rates of DNN and TDNN, enhancing their applicability in dynamic environments. Our theoretical analysis confirms that the statistical learning rates for the mean utility function and noise distribution are minimax optimal. We also derive a regret bound that illustrates the critical interaction between model dimensionality and noise distribution smoothness, deepening our understanding of dynamic pricing under varied market conditions. These contributions offer substantial theoretical insights and practical tools for implementing effective, data-driven pricing strategies, advancing the theoretical framework of pricing models and providing robust methodologies for navigating the complexities of modern markets.

3 - Uncertainty Estimation by Density Aware Evidential Deep Learning

TaeSeong Yoon, KAIST, Daejeon, Korea, Republic of, Heeyoung Kim

Evidential deep learning (EDL) has shown remarkable success in uncertainty estimation. However, there is still room for improvement, particularly in out-of-distribution (OOD) detection and classification tasks. The limited OOD detection performance of EDL arises from its inability to reflect the distance between the testing example and training data when quantifying uncertainty, while its limited classification performance stems from its parameterization of the concentration parameters. To address these limitations, we propose a novel method called Density Aware Evidential Deep Learning (DAEDL). DAEDL integrates the feature space density of the testing example with the output of EDL during the prediction stage, while using a novel parameterization that resolves the issues in the conventional parameterization. We prove that DAEDL enjoys a number of favorable theoretical properties. DAEDL demonstrates state-of-the-art performance across diverse downstream tasks related to uncertainty estimation and classification.

4 - High-Dimensional (Group) Adversarial Training in Linear Regression

Yiling Xie, Georgia Tech, Atlanta, GA, United States

Adversarial training can achieve robustness against adversarial perturbations and has been widely used in machine-learning models. This paper delivers a nonasymptotic consistency analysis of the adversarial training procedure under ℓ_∞ -perturbation in high-dimensional linear regression. It will be shown that, under the restricted eigenvalue condition, the associated convergence rate of prediction error can achieve the minimax rate up to a logarithmic factor in the high-dimensional linear regression on the class of sparse parameters. Additionally, the group adversarial training procedure is analyzed. Compared with classic adversarial training, it will be proved that the group adversarial training procedure enjoys a better prediction error upper bound under certain group-sparsity patterns.

5 - Optimal Classification-Based Anomaly Detection with Neural Networks: Theory and Practice

Tian-Yi Zhou, Georgia Institute of Technology, Atlanta, GA, United States

Anomaly detection is an important problem in many application areas, such as network security. Many deep learning methods for unsupervised anomaly detection produce good empirical performance but lack theoretical guarantees. By casting anomaly detection into a binary classification problem, we establish non-asymptotic upper bounds and a convergence rate on the excess risk on rectified linear unit (ReLU) neural networks trained on synthetic anomalies. Our convergence rate on the excess risk matches the minimax optimal rate in the literature. Furthermore, we provide lower and upper bounds on the number of synthetic anomalies that can attain this optimality. For practical implementation, we relax some conditions to improve the search for the empirical risk minimizer, which leads to competitive performance to other classification-based methods for anomaly detection. Overall, our work provides the first theoretical guarantees of unsupervised neural network-based anomaly detectors and empirical insights on how to design them well.

TD68

Building B Level 4 B408

Interpretable Machine Learning

Invited Session

Data Mining

Chair: Cynthia Rudin, Duke University, Durham, NC, United States

Co-Chair: Srikar Katta, Duke University, Durham, NC, United States

1 - Interpretable Neural Networks with Prototypes

Zhicheng Guo, Duke University, Durham, NC, United States

Interpretable machine learning holds transformative potential in high-stakes medical domains, where model transparency is critical for clinical adoption. In this talk, we present a prototypical part-based neural network designed to assist clinicians in classifying EEG patterns along the Ictal–Interictal–Injury continuum to prevent serious brain damage. The model leverages case-based reasoning by grounding decisions in clinically meaningful waveform prototypes, resulting in improved diagnostic accuracy and interoperability. However, developing high-performing interpretable models remains heavily constrained by the availability of machine learning expertise and the efficiency of collaboration between ML researchers and domain experts. To address this challenge, we introduce a framework for editing prototype-based models in real time using Rashomon sets - collections of diverse models with similarly high performance. This approach alleviates the interaction bottleneck by enabling users to explore and generate high-performance, personalized models that align with their domain-specific priors in just minutes, all without requiring a machine learning expert. By shifting control into the hands of clinicians and other domain experts, this Rashomon-based editing paradigm democratizes the creation of trustworthy, interpretable models across a variety of high-stakes medical applications. Together, these contributions advance human-centered AI in medicine, where the collaboration between models and experts is both intuitive and transparent.

2 - Robust Fine-Tuning with LLM Knowledge and Reasoning: News Headline Generation

Tong Wang, Yale University, New Haven, CT, United States, K. Sudhir, Hengguang Zhou

We address reward-hacking failures that arise when aligning language models through preference-based fine-tuning. Our method, **Knowledge-Regularized Alignment (KRA)**, augments each training prompt with concise, validated explanations of *why* one response is preferred over another. These explanations are first hypothesized by a large language model via abductive reasoning, then checked against data before being prepended to the prompt. The added “why-knowledge” guides the model toward behavior justified by coherent reasoning rather than superficial reward cues. Applied to abstractive headline generation, KRA improves both catchiness and relevance while markedly reducing reward-hacking artifacts. The framework is model-agnostic and complements existing techniques such as Direct Preference Optimization.

3 - Simultaneous Inference for Heterogeneous Treatment Effects in Observational Causal Inference

Srikar Katta, Duke University, Durham, NC, United States

While traditional causal inference methods answer whether a treatment works on average, modern causal questions require understanding *for whom* and *under what conditions* a treatment is effective. This shift toward personalized causal inference has led to the explosion of integrating flexible statistical and machine learning methods into the causal inference pipeline. However, as questions surrounding heterogeneity often explore a vast space of potential subgroups or covariate interactions, they implicitly give rise to a multiple hypothesis testing problem. Existing approaches typically either neglect this issue or apply simplistic corrections that are ill-suited to the adaptive nature of modern estimation procedures. Consequently, many subgroup findings risk being false discoveries, undermining the credibility of empirical claims. We present a solution to this problem, prove that our proposed estimator exhibits useful double robustness properties, enables simultaneous inference, and is end-to-end interpretable. Our simulations highlight the issues with existing approaches and demonstrate that our procedure is both more accurate and more flexible than alternatives.

4 - Explaining the Explanations

Xuefei Lu, SKEMA Business School, Suresnes, France, Emanuele Borgonovo, Cynthia Rudin, Elmar Plischke

Explainable artificial intelligence (XAI) is increasingly essential for regulatory compliance and model transparency. Concerns have arisen regarding the explanations generated by widely used methods such as SHAPs, LIME, or other permutation-based techniques. These issues include extrapolation or out-of-distribution problems inherent in many XAI approaches, where synthetic inputs used to calculate explanations may fall outside the data manifold or be impractical in real-world applications. Consequently, this may result in misleading or untrustworthy outcomes. We propose a systematic framework to evaluate the trustworthiness of popular XAI methods by incorporating both statistical and visual diagnostic techniques. Our methodology includes non-parametric tests of model outputs on original and synthetic predictions, as well as anomaly detection of less likely input combinations. We emphasize the importance of identifying these issues to ensure reliable and robust XAI results. The proposed methods are demonstrated using both analytical and real datasets.

TD69

Building B Level 4 B409

Constraint Learning: When a Neural Network Walks into an Optimization Model

Invited Session

Computing Society

Chair: Thiago Serra, University of Iowa, Iowa City, IA, United States

1 - Neural Heuristics for Mathematical Optimization via Value Function Approximation

Justin Dumouchelle, University of Toronto, Toronto, ON, Canada, Rahul Patel, Esther Julien, Jannis Kurtz, Merve Bodur, Elias B. Khalil

Mathematical optimization is an invaluable framework for modeling and solving decision-making problems with many successes in single-level deterministic problems (e.g., mixed-integer linear or nonlinear optimization). However, many real-world problems require accounting for uncertainty or the reaction of another agent. Paradigms such as stochastic programming, bilevel optimization, and robust optimization can model these situations but are much slower to solve than their deterministic counterparts, especially when discrete decisions must be made. This work demonstrates how a single framework based on value function approximation and optimizing over trained neural networks can be adapted to all three domains. Empirically, we find solutions of similar, and in some cases significantly better, quality than state-of-the-art algorithms in each field, often within a fraction of the running time. The datasets and three frameworks, Neur2SP (NeurIPS'22), Neur2RO (ICLR'24), and Neur2BiLO (NeurIPS'24), are open-sourced for further research.

2 - Optimization over Deep Relu Networks Using Penalty Methods

Anirudh Subramanyam, The Pennsylvania State University, University Park, PA, United States, Waquar Kaleem, Henri Lefebvre, Taehyeon Kwon

Optimization over trained ReLU networks is critical for tasks such as verification, compression, and embedding learned constraints into decision-making models. However, exact methods like MILP reformulations and cutting-plane approaches scale poorly with network depth and size, leading to large constraint systems and weak relaxations. We propose a Penalty Alternating Direction Method (PADM) that enforces ReLU complementarity through penalty terms in the objective, avoiding the use of binary variables. Under mild assumptions, we prove convergence to stationary points. Experiments on benchmark problems with very deep networks show that PADM quickly finds high-quality feasible solutions.

3 - Strategies for Solving Mixed Integer Linear Optimization Problems with Embedded Relu Artificial Neural Networks

Alexander Klimek, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany, Christoph Plate, Mirko Hahn, Caroline Ganzer, Kai Sundmacher, Sebastian Sager

When artificial neural networks (ANNs) are formulated with the ReLU activation function, they can be integrated into mixed integer linear programming problems (MILP) using a big-M reformulation. This enables optimization over trained surrogate models using established MILP methods. However, these optimization tasks appear to become challenging for increasing number of neurons and layers, and for optimization problems with multiple ANNs embedded. One reason are the big-M coefficients arising in the reformulation which grow exponentially with the number of layers.

In this work, we examine four ways to speed up MILP optimization over trained ReLU ANNs:

1. Optimization-based bound tightening (OBBT)
2. Novel *a posteriori* ReLU scaling
3. Clipped ReLU
4. Regularization

Our results suggest that OBBT is effective for all trained networks, reducing big-M coefficients by half. The ReLU scaling or clipped ReLU decrease big-M coefficients to a similar degree. Regularization delivers the highest reduction of the big-M coefficients by up to two orders of magnitude. However, only OBBT and regularization appear successful in lowering computational time significantly, to around half, and to as low as 3% of the original time, respectively. This implies that reducing big-M coefficients alone does not suffice. Analysis of the number of linear regions and the number of stable neurons reveals the effectiveness of regularization in improving both, while OBBT does not incur changes in either. Importantly, regularization greatly raises the number of instances that can be solved at all within the specified time limit.

4 - Optimization over Trained Neural Networks: Difference-of-Convex Algorithm and Application to Data Center Scheduling

Xinwei Liu, University of Michigan, Ann Arbor, MI, United States, Vladimir Dvorkin

When solving decision-making problems with mathematical optimization, some constraints or objectives may lack analytic expressions but can be approximated from data. When an approximation is made by neural networks, the underlying problem becomes optimization over trained neural networks. Despite recent improvements with cutting planes, relaxations, and heuristics, the problem remains difficult to solve in practice. We propose a new solution based on a bilinear problem reformulation that penalizes ReLU constraints in the objective function. This reformulation makes the problem amenable to efficient difference-of-convex algorithms (DCA), for which we propose a principled approach to penalty selection that facilitates convergence to stationary points of the original problem. We apply the DCA to the problem of the least-cost allocation of data center electricity demand in a power grid, reporting significant savings in congested cases.

5 - Optimization over Trained (and Sparse) Neural Networks: a Surrogate Within a Surrogate

Thiago Serra, University of Iowa, Iowa City, IA, United States, Hung Pham, Aiden Ren, Ibrahim Tahir, Jiatai Tong

We can approximate a constraint or an objective function that is uncertain or nonlinear with a neural network that we embed in the optimization model. This approach, which is known as constraint learning, faces the challenge that optimization models with neural network surrogates are harder to solve. Such difficulties have motivated studies on model reformulation, specialized optimization algorithms, and—to a lesser extent—pruning of the embedded networks. In this work, we double down on the use of surrogates by applying network pruning to produce a surrogate of the neural network itself. In the context of using a Mixed-Integer Linear Programming (MILP) solver to verify neural networks, we obtained faster adversarial perturbations for dense neural networks by using sparse surrogates, especially—and surprisingly—if not taking the time to finetune the sparse network to make up for the loss in accuracy. In other words, we show that a model with bad classification performance can still be a good—and more efficient—optimization surrogate.

Tuesday, October 28, 4:15 PM - 5:30 PM

TE01

Building A Level 3 A301

SAS

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Building and Solving Optimization Models with Sas

Rob Pratt, SAS, Cary, NC, United States

SAS offers extensive analytic capabilities, including machine learning, deep learning, natural language processing, statistical analysis, optimization, and simulation. SAS analytic functionality is also available through the open, cloud-enabled design of SAS® Viya®. You can program in SAS or in other languages – Python, Lua, Java, and R. OPTMODEL from SAS provides a powerful and intuitive algebraic optimization modeling language and unified support for building and solving LP, MILP, QP, conic,

NLP, constraint programming, network-oriented, and black-box models. This showcase will include an overview of the optimization capabilities and demonstrate recently added features.

TE03

Building A Level 3 A303

Data-Driven Modeling and Decision-Making for System Prediction and Monitoring

Invited Session

Quality, Statistics and Reliability

Chair: Dongmin Li, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Xiaochen Xian, Georgia Institute of Technology, Atlanta, GA, United States

1 - Dimension Reduction via Variable Selection in Importance Sampling

Chenfei Li, University of Michigan, Ann Arbor, MI, United States, Jaeshin Park, Eunshin Byon

Importance sampling is a powerful variance reduction technique used to minimize uncertainty in stochastic simulation. However, it often faces the curse of dimensionality, as the input dimension grows. Grounded in the parsimony principle, we introduce a variable selection metric, which helps reduce input dimension. Additionally, we devise a new cross-validation procedure which accounts for unique challenges present within the importance sampling context. Our method balances concentrating on key variables while exploring broader input areas, thereby enhancing the robustness of estimated outcomes. While broadly applicable, we demonstrate its effectiveness in a nonparametric importance sampling setting.

2 - Wavelet Gaussian Process for Multi-Scale Monitoring of Imaging Data

Runsang Liu, The Pennsylvania State University, University Park, PA, United States, Hui Yang

Advanced sensing brings a plethora of imaging data to handle process variations in additive manufacturing (AM). Realizing the full potential of sensor-based statistical process control (SPC) methods greatly depends on the delineation of defect characteristics and anomalous patterns. However, traditional methods primarily focus on analyzing and monitoring AM images in a single scale. It is common that these images contain hidden patterns pertinent to process deviations that are not accessible in a single-scale view. This paper presents a new wavelet Gaussian process (WGP) framework for multi-scale analysis of variance (ANOVA) and image-guided SPC of AM. First, AM imaging data are examined through multi-scale wavelet lenses to reveal hidden defects in different spatial-frequency resolutions. Second, WGP models are developed to analyze different variance components in wavelet coefficients. Specifically, we separate the variance contribution of stochastic variation, spatial-temporal effects, and process shifts. We further design generalized likelihood ratio (GLR) formulation and multi-scale hypothesis testing approaches to fuse test statistics from different wavelet sub-bands for AM quality monitoring. The proposed WGP approach is evaluated and validated through simulation and real-world case studies. Experimental results demonstrated the effectiveness of the proposed WGP method for multi-scale analysis and monitoring of imaging data.

3 - State Space Model with Random Effects for Autocorrelated Multivariate Ordinal Data

Zengchenghao Xia, University of Iowa, Iowa City, IA, United States

Multivariate ordinal data, frequently manifested as Likert-scale responses, are ubiquitous in healthcare evaluations and fatigue assessment where items are collected repeatedly from numerous subjects. These intensive longitudinal data exhibit both cross-sectional correlations between items and temporal autocorrelations within subjects. However, current statistical methodologies often lack frameworks specifically designed to systematically model the crucial multivariate dependencies inherent in such ordinal longitudinal data structures.

To address this limitation, we propose a state-space modeling framework that systematically accommodates both the ordinal nature of multivariate Likert-scale responses and inter-individual variability via subject-specific random effects. The underlying latent states are posited to follow vector auto-regressive (VAR) dynamics, a specification inherently accommodating flexible contemporaneous and lagged correlation structures. Within the observation model, these latent states then determine the ordinal response for each Likert item relative to a set of ordered thresholds, which themselves evolve dynamically via random walks to capture temporal shifts in response patterns. Parameter estimation, employing a Monte Carlo Expectation-Maximization (MCEM) algorithm with integrated particle filtering and smoothing techniques, yields satisfactory estimators for both model parameters and latent process variables. The practical utility and advantages of this approach are confirmed through numerical simulations and a case study, which highlights its accurate parameter estimation and outperformance of traditional benchmarks in accurately representing the complex interdependencies of multivariate ordinal responses.

4 - On-Demand Machine Learning for Resource-Constrained Classification

Xiaochen Xian, Georgia Institute of Technology, Atlanta, GA, United States, Dongmin Li, Minsung Kang, Gonen Singer, Hongcheng Liu, Md Mahmudul Hasan

In many real-world applications, classification tasks are subject to resource constraints, such as limited resources and budgetary restrictions, which limit the number of testing samples that can be classified into a particular label. While existing methods integrate prediction and optimization models to enhance classification performance, there is a lack of a general approach that incorporates the constraints into the learning process to optimize misclassification cost while ensuring constraint in compliance on testing data. To address this gap, we introduce a novel machine learning framework, “on-demand learning (OL),” which adapts label assignments on testing data based on resource constraints to minimize misclassification cost. OL incorporates cost-sensitive learning techniques, introducing a dynamic class weight into the loss function to integrate prediction and optimization. Leveraging simulated annealing, OL iteratively adjusts the class weight based on prediction results and constraints to identify the optimal classifier. This constraint-dependent adjustment makes the learned model adaptive, adjusting feature prioritization based on constraints, which enables flexible and efficient resource allocation under real-world constraints and costs. Experiments on synthetic and real-world datasets demonstrate the adaptability of OL and its superiority over benchmark methods across multiple classification metrics, highlighting the benefits of incorporating resource constraints into the learning process.

TE04

Building A Level 3 A304

Data-Driven Discovery: Integrating Statistics and AI for Scientific Insights

Invited Session

Quality, Statistics and Reliability

Chair: Ridwan Olabiyi, Arizona State university, Tempe, AZ, 85281, United States

Co-Chair: Yuhao Zhong, Virginia Tech, Blacksburg, VA, 77840, United States

1 - Uncertainty Aware Data Driven Discovery of Governing Equations

Yuhao Zhong, Virginia Tech, Blacksburg, VA, United States

In complex physical systems where domain knowledge is limited or undeveloped, dimensional analysis serves as a powerful tool for characterizing system behavior and discovering governing equations from observed data. However, traditional approaches face two key limitations. First, they often produce non unique dimensionless groups and rely on ad-hoc strategies to infer governing laws. Second, in the absence of comprehensive domain knowledge, these approaches lack mechanisms to handle data uncertainty, making it difficult to quantify the reliability or confidence of the resulting equations. To address these challenges, we propose a two-stage framework that blends projection pursuit Gaussian process (GP) regression with dimensional analysis. In the first stage, we identify physically invariant combinations of input variables. In the second stage, we model the system’s response using a GP over the projected dimensionless space, enabling uncertainty quantification through the GP’s posterior distribution.

2 - Candice: Causal Discovery of Nonlinear Dynamics Through Counterfactual Explanations

Ridwan Olabiyi, Arizona State university, Tempe, AZ, United States

Discovering governing equations from data is a core challenge in scientific research, particularly for complex systems where first-principles derivations are impractical. Recently, sparse learning techniques have been used to extract physics from high-dimensional, noisy data. However, the recovered equations either include extraneous terms or miss critical ones, limiting their generalizability. To overcome this, we propose CaNDiCE—Causal Discovery of Nonlinear Dynamics through Counterfactual Explanations—a parsimonious learning approach to recover the governing equations by identifying causal physics terms. Counterfactuals are generated by minimally perturbing the parameters of the candidate governing equations such that the perturbed system behavior is out-of-distribution with respect to the observed data. The physics terms that drive this transition are deemed causal. We demonstrate the effectiveness of CaNDiCE on benchmark systems, including the Lotka–Volterra and Lorenz systems, where it eliminates extraneous terms while closely matching ground-truth dynamics, outperforming baseline methods by roughly 93% in terms of mean squared error. Our results show that counterfactual explanations serve as a powerful approach for physics discovery, and the proposed methodology is readily extendable to other data-driven applications, enhancing both trust and interpretability.

3 - Machine Learning Powered Graph Network Analysis for Solid-State Battery Cathodes

Zebin Li, University of Wisconsin-Madison, Madison, WI, United States, Shimao Deng, Yijin Liu, Jiamian Hu

The microstructure of particulate composites, particularly the phase boundaries and interfacial contacts, can critically affect the system performance. However, the processes and mechanisms that govern the relationship between microstructure, property, and performance remain highly complex and insufficiently understood. Thus, developing deep insights of how microstructural features impact performance is crucial. Here, we take the composite cathode of SSBs as an example, where triple phase boundaries (TPBs) between particles of mixed active cathodes (e.g., $\text{LiNi}_x\text{Mn}_y\text{Co}_z\text{O}_2$, $x + y + z = 1$, denoted as NMC), solid-state electrolytes (SSE), and conductive carbons (e.g., graphite) play a crucial role in battery performances. X-ray microscopy is a powerful non-destructive imaging technique that enables high-resolution analysis of microstructures. However, it is a computational challenge to process and analyze the high-dimensional and large-volume data with high accuracy, automation, efficiency. In this work, we propose an automated phase segmentation framework supporting particle network analysis for SSBs cathodes based on deep learning. In particular, three components in cathodes, i.e., NMC, SSE, and graphite, are first segmented into three phases using U-Net. Then, graphs are constructed based on the segmentation maps to reflect the connections among different phases. Furthermore, according to the constructed graphs, the TPBs can be searched in a quick and automatic way, enabling the functional relevant analysis for the cathode of SSBs. Our work not only provides an automatic pipeline enabling the analysis from X-ray images to TPBs of SSBs, but also reveals the electrochemical properties of TPBs for better understanding of the mechanism of SSBs.

TE05

Building A Level 3 A305

Decision-Focused Learning

Invited Session

Decision Analysis Society

Chair: Mo Liu, University of North Carolina-Chapel Hill, Hanes 403, Chapel Hill, NC, 27599, United States

1 - Feature-Dependent Marginal Values of One Data Point: Active Label Acquisition in Assortment Optimization

Mo Liu, University of North Carolina-Chapel Hill, Chapel Hill, NC, United States, Junyu Cao, Zuo-Jun Shen

Predicting customers' preferences based on their features is crucial for personalized assortment optimization. When building this prediction model, using informative data can significantly increase the expected revenue from personalized assortments. This paper studies how to sequentially collect informative data to construct this prediction model. We introduce a novel concept, the 'value of one data point,' which evaluates the marginal contribution of acquiring a specific customer's preference to the expected revenue in personalized assortment optimization, given the existing training set. Notably, this value drops to zero once the optimal assortment for this specific customer is determined. To estimate this value and identify important customers for acquiring their preferences, we derive a feature-dependent upper bound. This bound provides significant insights into the importance of each data point for revenue growth. Based on this upper bound, we develop a personalized incentive policy for effectively collecting survey data from customers to obtain their preferences. We provide non-asymptotic guarantees for both the cumulative incentives and the revenue from the final prediction model. Theoretically, we show that our personalized incentive policy requires smaller cumulative incentives than any fixed incentive policy to achieve the same level of revenue. Additionally, our numerical experiments with real-world and synthetic datasets validate the effectiveness of our personalized incentive algorithms over fixed strategies.

2 - More Data or Better Data? Impact of Costly Data Collection on the Newsvendor Problem

Zijin Zhang, Carroll School of Management, Boston College, Chestnut Hill, MA, United States, Hyun-Soo Ahn, Lennart Baardman

Data-driven algorithms and policies have flourished in many areas of operations management. While many applications can utilize preexisting data, there are many cases where collecting more and better data is cost-prohibitive and time-consuming. Previous research has mainly focused on improving decision-making for given data, leaving a gap in understanding the optimal quantity and quality of data needed. Our paper studies how the volume (quantity) and noise (quality) of a dataset jointly affects the outcome of data-driven decision-making in the newsvendor problem. We show that, when noise is present, the traditional data-driven methods can result in mediocre outcomes regardless of how large the dataset is. To mitigate the effect of noise, we propose a novel policy that denoises the dataset in order to minimize the noise-robust in-sample cost. Using this denoised approach, we analyze the retailer's out-of-sample newsvendor cost as a function of sample size and noise scale. Our analysis quantifies the marginal newsvendor cost on each dimension, and we find that the newsvendor outcome is non-monotone in both the quantity and quality dimensions. Furthermore, given the cost of data collection, we develop a series of data collection policies and show that collecting a dataset of well-balanced quantity and quality is important.

3 - End-to-End Optimization Framework for Selective Decision-Making

Michael Huang, CUNY Baruch College, New York, NY, United States, Bradley Rava

Modern Artificial Intelligence (AI) tools have reached a point where policymakers can potentially engage in fast and autonomous large-scale decision-making in high-risk settings. However, despite advancements, the adoption of AI in these settings remains controversial as they can be inaccurate and disproportionately affect minority or underrepresented groups and individuals. As a result, many policymakers revert to human-based decision-making, which, unfortunately, can be equally ineffective as decisions are often delayed or require high labor costs. To obtain the best of both worlds, we propose a synergistic Human-AI decision-making framework that efficiently splits decision-making between humans and AI. The framework should rigorously control the trade-off between human decision-making cost vs. AI decision-making cost and be flexible enough to accommodate fairness considerations across different groups.

4 - Learning to Price with Resource Constraints: from Full Information to Machine-Learned Prices

Ruicheng Ao, MIT, Cambridge, MA, United States, Jiashuo Jiang, David Simchi-Levi

We study the dynamic pricing problem with knapsack, addressing the challenge of balancing exploration and exploitation under resource constraints. We introduce three algorithms tailored to different informational settings: a Boundary Attracted Re-solve Method for full information, an online learning algorithm for scenarios with no prior information, and an estimate-then-select re-solve algorithm that leverages machine-learned informed prices with known upper bound of estimation errors. The Boundary Attracted Re-solve Method achieves logarithmic regret without requiring the non-degeneracy condition, while the online learning algorithm attains an optimal $\mathcal{O}(\sqrt{T})$ regret. Our estimate-then-select approach bridges the gap between these settings, providing improved regret bounds when reliable offline data is available. Numerical experiments validate the effectiveness and robustness of our algorithms across various scenarios. This work advances the understanding of online resource allocation and dynamic pricing, offering practical solutions adaptable to different informational structures.

5 - Dissecting the Impacts of Model Misspecification in Data-Driven Optimization

Haixiang Lan, Columbia University, New York, NY, United States, Adam Elmachtoub, Henry Lam, Haofeng Zhang

Data-driven optimization aims to translate a machine learning model into decision-making by optimizing decisions on estimated costs. Such a pipeline can be conducted by fitting a distributional model which is then plugged into the target optimization problem. While this fitting can utilize traditional methods such as maximum likelihood, a more recent approach uses estimation-optimization integration that minimizes decision error instead of estimation error. Although intuitive, the statistical benefit of the latter approach is not well understood yet is important to guide the prescriptive usage of machine learning. In this paper, we dissect the performance comparisons between these approaches in terms of the amount of model misspecification. In particular, we show how the integrated approach offers a "universal double benefit" on the top two dominating terms of regret when the underlying model is misspecified, while the traditional approach can be advantageous when the model is nearly well-specified. Our comparison is powered by finite-sample tail regret bounds that are derived via new higher-order expansions of regrets and the leveraging of a recent Berry-Esseen theorem.

TE06

Building A Level 3 A306

Consumer Psychology and Digital Engagement in E-Commerce

Contributed Session

Contributed

Chair: Shuzhang Cai, The University of Texas at Dallas, Richardson, TX, United States

1 - Why Full Refunds Prevail: A Product Fitness Perspective

Chuanshuai Ru, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Fangruo Chen, Weitao Ren, Wenqiang Xiao

This paper investigates the design of return policies in online markets characterized by consumer heterogeneity, product misfit risk and information asymmetry. We consider a monopolistic seller serving a dual-segment market, consisting of informed and uninformed consumers. Departing from prior studies that advocate cost-based partial-refund strategies, we identify product fitness, rather than cost, as the central driver of return policy design. Our analysis reveals that the optimal policy can be implemented via a full-refund contract, achieving operational simplicity without sacrificing profitability. We also find that when mismatched products have residual values, the seller may shift to lower selling prices and partial (or no) refunds to encourage product retention and minimize returns, especially when uninformed consumers dominate the market. The analysis further reveals that full-refund policies are more appropriate for low-value goods, while partial-refund policies better fit high-value products. Our findings offer a unified framework linking product fitness and return strategies, delivering both theoretical contributions and practical insights for online sellers managing returns.

2 - Digital Nudging for Display Optimization in Online Product Listing Pages

Alara Tascioglu, ESCP Business School, Paris, France, Gurhan Kok, Selcuk Karabati

Problem definition: As digital commerce continues to expand, retailers must optimize their Product Listing Pages (PLPs) to guide consumer attention and maximize profitability. While prior research has largely focused on product rankings, the influence of product display on a PLP (location and size) on consumer choice remains underexplored.

Methodology/results: We introduce a location-based Multinomial Logit (LMNL) model for customer choice and jointly optimize default rankings and location sizes to enhance expected profit margins per customer. We also develop a computationally efficient heuristic that closely approximates the model.

Through three structured lab experiments, we find that consumer behavior aligns with LMNL, and varies by device: desktop users exhibit a centripetal U-shaped viewing pattern with faster diminishing returns to space, while mobile users focus more on later scrolls. We optimize PLPs based on parameters estimated in the experiments and verify through additional experiments that these optimized PLPs increase expected profit by 15.13% compared to PLPs with identical location sizes, while randomly selected large displays and products lower them by 18.17%.

Managerial implications: These findings highlight the limitations of test-and-learn approach (A/B testing) due to the combinatorial nature of PLP design and the need for a model-based and data-driven approach. We provide a structured framework for parameter estimation and scalable PLP optimization across devices and PLP structures. By bridging digital nudging, choice modeling, online display optimization, and experimental analysis, our research improves interface design and customer experience.

3 - The Impact of Positive Review Reward on Consumer Review and Product Return: Effect of Reference Point

Huimin Wang, Tianjin University, Tianjin, China, People's Republic of, Yize Li, Min Cai

Online reviews shapes consumer purchase decisions. To encourage positive review, many sellers implement the positive review rewards strategy (PRRS). While PRRS may compromise review authenticity despite regulatory restrictions, their actual impact remains unclear. Drawing on reference point theory, this study empirically examines how PRRS influences consumer reviewing and product returns. Analyzing survey data from 634 consumers in China, we find that PRRS effectively increases genuine positive reviews without significantly boosting fake ones. However, the surge in positive reviews raises consumers' reference point of the product value, leading to subsequent dissatisfaction and negative reviews that counteract the initial boost in purchasing and product returns. Overall, PRRS enhances product sales and consumer surplus. Moreover, our findings indicate low-income consumers in the early stages of their careers are more susceptible to this strategy, making them more prone to posting fake reviews.

4 - From Browsing to Purchasing: Exploring the Influence of User Comments on the Desire Infiltration of Short-Video Sales

Tong Li, Guangdong University of Foreign Studies, Guangzhou, China, China, People's Republic of

With the growing prevalence of short-video e-commerce, consumers experience a complex and dynamic information perception process as they transition from interest-driven browsing to emotion-driven purchasing behavior. Based on the Desire Infiltration Theory, this study explores how information perception in comment sections can guide users to form purchase intentions, aiming to address marketing conversion challenges in the era of interest-based e-commerce and social commerce. Using a dataset from the Douyin platform, we obtained a dataset of 28341 product-promoting short videos and their corresponding 236870 comments, and analyzed the impact of the degree of consumer purchase confidence represented in comments on short-video sales. As each comment in short videos receives a different number of likes, and the number of likes will affect the order of comments in the comment section, we found that 80% of the cumulative total likes came from 20% of the comments, confirming the validity of Pareto's Law. Further empirical testing revealed that the consumer purchase confidence reflected in the top 20% of highly liked comments has a positive impact on short-video sales, while the purchase confidence reflected in all comments shows no significant effect on sales. Our research confirms that in the era of fragmented information in short videos, users under fast-paced information browsing patterns do not have the patience to read all information completely. Therefore, short-video e-commerce platforms need to effectively understand the key influences of users' information perception during the desire infiltration process, so as to formulate effective marketing strategies based on key information perception.

5 - Review Vs Tips: Which One is More Influential? Investigating the Impact of Yelp Tips on Restaurant Reputation

Aindrila Chakraborty, Metro State University, St Paul, MN, United States, Jaymeen Shah

Online recommendation platform Yelp significantly shapes consumer perceptions and informed decisions about local businesses through reviews and Tips. While Reviews mainly focus on detailed user experiences with a business, Tips are more concise, significantly shorter than reviews, and primarily focus on providing immediate, specific insights and recommendations, especially to mobile users. An extensive body of literature exists about the different aspects of reviews and their impact. However, Yelp Tips, which provides concise and more time-sensitive business recommendations, has yet to be explored in the literature. In this study, we aim to bridge this gap by examining the restaurant businesses and investigating whether Yelp Tips, apart from reviews, have any effect on overall restaurant reputation. Utilizing a text analytics-based approach combined with feature engineering on a nationwide dataset of restaurants on Yelp, this study provides significant insights into the distinct role of Tips in shaping customer expectations. Results suggest a substantial impact of Tips on time-sensitive user decisions by forming a quick, real-time, immediate reputation of restaurants to time-sensitive users, especially if they are traveling and in a new place. The insights from this study will enable business owners to make informed decisions about critical business functions to improve customer satisfaction and to manage their online reputation more effectively.

6 - Explainable Recommendations with Personalized Review Summaries

Shuzhang Cai, The University of Texas at Dallas, Richardson, TX, United States, Shaojie Tang

Online platforms such as TripAdvisor now provide AI-generated summaries to condense user reviews for hotels, restaurants, tourist attractions, etc. While these summaries help users quickly acquire a general overview, they often lack personalization and fail to align with individual user preferences across different aspects. In this work, we propose a novel framework that integrates contrastive learning and graph learning to capture nuanced user-item interactions and deliver explainable recommendations in the form of personalized summaries.

TE07

Building A Level 3 A307

Next-Gen AI, Quantum, and Cognitive Technologies for Decision-Making

Contributed Session

Contributed

Chair: Yuchen Liang, Tsinghua University, Beijing, 100084

1 - Automating Attribute-Based Access Control with LLM Distillation and Graph-Based Policy Modeling

Mian Yang, Rutgers Business School, Newark, NJ, United States

Modern enterprises require scalable and explainable access control systems to enforce security and compliance. However, deriving fine-grained Attribute-Based Access Control (ABAC) rules from policy documents written in natural language remains a significant challenge. In this study, we propose an end-to-end framework for automating ABAC rule extraction and enforcement using Large Language Models (LLMs), knowledge distillation, and graph-based representations.

Our first contribution is a novel flow graph model that encodes ABAC rules as directed graphs with support for permit/deny decisions, conflict detection, and redundancy checking. These graphs enable efficient enforcement and policy reasoning via graph database queries.

Our second contribution introduces a distilled LLM pipeline that identifies subject, object, and action components from natural language access control policies (NLACPs), transforming unstructured sentences into structured ABAC rules. By combining distillation techniques with fine-tuned BERT models, we significantly reduce model complexity while preserving high extraction accuracy.

Together, our approach bridges the gap between natural language policy documents and machine-enforceable access control systems. Experimental results demonstrate the effectiveness of our framework in automating policy extraction, ensuring rule correctness, and enabling scalable enforcement via Neo4j-based graph querying.

2 - Automating Invoice Processing with Rpa and LLM Knowledge Distillation

Yang Li, The George Washington University, Washington DC, DC, United States, Mian Yang

Traditional business processes such as invoice handling remain time-consuming and error-prone in enterprise financial operations. This study presents an end-to-end automated invoice processing pipeline that integrates Robotic Process Automation (RPA), Optical Character Recognition (OCR), and fine-tuned BERT-based models to extract key information from PDF and JPG invoice documents. To improve model efficiency and enable practical deployment, we introduce a knowledge distillation framework that transfers knowledge from a large pretrained language model to a lightweight student model. This allows for faster inference while preserving high extraction accuracy. Extracted data is automatically validated and integrated into enterprise resource planning (ERP) systems through RPA bots. The proposed solution is scalable, cost-effective, and capable of reducing manual workload in real-world financial workflows. Experimental results demonstrate strong generalizability and minimal need for human intervention across diverse invoice formats.

3 - Generating Controllable Visual Stimuli Using Stable Diffusion: a Guide for Behavioral Experimental Research

Joann Liu, University of Florida, Gainesville, FL, United States, Jia Liu, Xinyi Sun

Visual stimuli, such as human faces and product presentation, are pivotal in studying human perception, judgment, and decision-making. Traditional methods for creating experimental stimuli involve significant challenges, including labor-intensive workflows, limited scalability and generalizability, and difficulties in isolating specific visual features. This paper bridges the gap between AI innovation and behavioral research by introducing Stable Diffusion that can generate high-fidelity, controllable visual stimuli in a scalable but cost-effective manner. Stable Diffusion also holds the potential to provide visually interpretable constructs by fine-tuning annotated image datasets. These capabilities can empower researchers to unlock great research opportunities on visual perception and decision-making. This paper starts by explaining the key concepts and techniques, followed by two sets of applications of Stable Diffusion to behavioral experimental research. These applications include systematic manipulations of facial features (e.g., smile intensity, gender neutrality, age progression, and beauty effort) and of product-related images (e.g., visual complexity, packaging design, and background setting), some of which are replications of the stimuli used in the existing literature. The paper further outlines future research directions for integrating generative AI with behavioral science, neuroscience, and adaptive experimental paradigms.

4 - Solving Multi-Stage Stochastic QUBOs with QAOA

Anthony Wilkie, University of Tennessee-Knoxville, Knoxville, TN, United States, Rebekah Herrman, Jim Ostrowski

Multi-stage stochastic optimization models sequential decision-making under uncertainty, where decisions at stage t are made prior to the realization of uncertainty in stage $t+1:t+T$. In many formulations, uncertainty is represented by discrete scenarios evolving between stages, leading to an exponential increase in problem size as the number of stages and scenarios grows. This poses a major challenge for classical solution techniques, particularly for binary decision problems. Quantum computing offers a promising framework for tackling such problems, especially in the domain of combinatorial optimization. The Quantum Approximate Optimization Algorithm (QAOA) is one such variational quantum algorithm designed to solve discrete optimization problems on near-term quantum devices. Extending our previous work, we propose two circuit encoding strategies for solving multi-stage stochastic quadratic unconstrained binary optimization (SQUBO) problems using QAOA. The first approach encodes the deterministic equivalent (DE) formulation by introducing separate decision variables for each scenario. The second employs scenario-qubits (SQ) to compactly represent uncertainty, using conditional unitaries to apply stage- and scenario-specific cost and mixing terms. We compare the performance of QAOA under both encodings in identifying optimal stage-wise decisions, and analyze the associated quantum resource requirements, including qubit counts and circuit complexity.

5 - Learning and Optimization for Low-Altitude Decision-Making: a Resource Allocation Framework Under Self-Imposed Information Constraints

Yuchen Liang, Tsinghua University, Beijing, China, People's Republic of, Chen Wang, Lefei Li

Long-range remote sensing missions often face predictable, intermittent communication with a central coordinator. In such cases, decision-making agents may operate without real-time connectivity, creating a decision setting with self-imposed information constraints. This setting poses two linked challenges: maintaining autonomous performance with limited information and allocating scarce computational resources. We develop a resource allocation framework that balances centralized guidance and local autonomy under these conditions, capturing trade-offs among information acquisition, resource allocation, and task performance. By abstracting physical details into decision-relevant parameters, the study provides a general formulation applicable to a range of intermittent-information problems.

TE08

Building A Level 3 A310

Harnessing AI for a Responsible Future

Invited Session

Information Systems

Chair: Kofi Arhin, Lehigh University, Bethlehem, PA, 18015, United States

1 - Generative AI Review Summaries: Cognitive and Behavioral Implications for Online Purchase Environments**Akshat Lakhiwal, University of Georgia - Terry College of Business, Bloomington, IN, United States**

Generative artificial intelligence review summaries (GARS) are increasingly deployed on online platforms to condense vast volumes of peer-generated feedback aimed at enhancing product evaluation; however, their current implementations may not align with traditional evaluative cues such as star ratings and detailed user comments. Drawing on prior research on mixed emotions and decision-making in online environments, this study theorizes that the congruence between GARS and conventional representations is vital for coherent information processing. Employing a multi-method experimental design that integrates eye-tracking, fNIRS, and behavioral analyses, we assess the effects of GARS-representation alignment on visual attention, neural activation, and purchase behavior. Our findings reveal that inconsistencies between AI-generated and user-generated content disrupt attentional focus, elevate cognitive effort, and reduce purchase likelihood, thereby underscoring the imperative for robust platform governance in GARS implementations. We discuss implications for platforms, including the need for regular audits of review synthesis processes and interventions to ensure representational congruence, thereby sustaining informed decision-making and preserving the integrity of digital marketplaces.

2 - A Bias-Aware Evaluation of AI Skill Recommendation Systems in Online Labor Markets**Brittany Green, University of Louisville, Louisville, KY, United States**

Online labor platforms increasingly use AI-based systems to recommend skills and guide freelancer development. While these systems offer personalization and efficiency, they may unintentionally interact with well-documented employer bias, leading to unequal and potentially harmful outcomes for underrepresented groups. Prior research has demonstrated that algorithmic systems can reinforce bias over time, yet the fairness implications of skill recommendation systems remain understudied. This study investigates how these systems promote or constrain advancement for groups disproportionately affected by employer bias. We model several skill recommendation strategies and evaluate their long-term impact on freelancer outcomes, including earnings and reputational trajectories. We parameterize a simulation model with hiring data from a major online labor market to assess how these strategies perform under varying levels of employer bias and to quantify disparities in outcomes across freelancer groups. Our findings contribute to the ethical design and governance of AI recommender systems by showing how personalization may mitigate or exacerbate systemic inequities in online labor markets.

3 - Measuring Social and Equity Impacts of LLMs to Assist Staff in Social Service Agencies**Jennah Gosciak, Cornell University, Ithaca, NY, United States**

We aim to evaluate efficacy-bias tradeoffs in AI systems as applied to public services, specifically through a LLM rollout (run by a non-profit partner) aimed to reduce administrative burdens in programs like SNAP (formerly known as food stamps). We first use representative data about statewide SNAP program rules to generate a benchmark dataset of common questions that a social service worker might receive from a SNAP client. We then perform a randomized experiment comparing accuracy in question-answering among (a) social workers, (b) social workers who can see LLM responses to the benchmark questions, and (c) LLMs themselves. In the experiment, we vary the accuracy of the LLM responses shown to different survey participants in order to observe the degree to which social workers may over-rely on LLM responses (even when they are incorrect). We also vary attributes (such as age, disability status, household citizenship status) to observe whether social workers or LLMs disproportionately answer questions incorrectly for certain potential SNAP clients. Finally, we calculate disparities in accuracy across dimensions related to biases, as well as measure changes in time-based efficiency and survey-based administrative burdens when LLMs are provided as a resource. Findings will have significant impacts on LLM procurement in government services; we aim to study whether balance can be achieved between easing administrative burdens while also improving benefits access for all clients.

4 - Shaping the Moral Boundaries of AI Systems: a Regulatory Focus Perspective of Hate Speech Detection by Generative Language Models**Kofi Arhin, Lehigh University, Bethlehem, PA, United States**

Automated content moderation systems rely on human-labeled data to train models that detect and flag harmful content online. Much of the existing literature has focused on improving inter-annotator agreement and intra-annotator consistency through clearer guidelines or better training. Yet, the task of labeling hate or offensive speech is not objective; it is shaped by the moral boundaries of annotators—boundaries that determine what is considered offensive, harmful, or acceptable. As such, existing efforts overlook the question of why annotators disagree in the first place. This question becomes even more pressing as generative AI (genAI) systems assume annotation responsibilities, either by supporting or substituting human labelers. It is therefore important to examine how the labeling decisions of genAI systems reflect context-sensitive moral reasoning similar to that of people. In this study, we examine how annotation guidelines, grounded in the regulatory focus theory, shape the labeling behavior of three genAI systems. We prompt each model with promotion-focused, prevention-focused, and neutral guidelines to classify 450 text samples each as offensive or not offensive. We find that both cross-model variability is influenced by how annotation guidelines are framed, and that these effects are moderated by the degree of linguistic ambiguity in offensive messages and target group saliency. We theorize these effects as evidence of AI moral drift to explain the malleability of genAI moral judgment in response to motivational framing and linguistic ambiguity. Our findings contribute to existing and emerging work on fairness and reliability in AI-assisted content moderation.

TE09

Building A Level 3 A311

FinTech in Information Systems: Blockchain, Decentralized Organizations and GenAI

Invited Session

Information Systems

Chair: Ye Liu, University of Washington, Seattle, United States

1 - Distributed Insurance**Baikun Leng, University of Washington Seattle, Seattle, WA, United States, Rowena Gan**

In the traditional insurance industry, the evaluation of claims is often biased due to the inherent interest of claim assessors in maximizing company profits. We explore the potential of decentralized insurance, powered by blockchain technology, as an alternative approach to mitigate such biases. In decentralized insurance models, claims are evaluated by a "court" of assessors who are independent of the insurance company. These assessors vote collectively to determine the outcome of a claim, with their compensation tied to their alignment with the majority decision. This mechanism incentivizes assessors to provide judgments that reflect the consensus of the group, rather than the financial interests of the insurance company.

2 - Optimal Arbitrage Timing Between Decentralized and Centralized Exchanges**Gustavo Grivol, New York University, NEW YORK, NY, United States, Josh Reed**

In decentralized finance, the timing in which blocks are released on a blockchain is influenced by the existence of arbitrage opportunities. We study a single agent arbitraging a risky asset between a decentralized exchange and a centralized exchange. The decentralized exchange is defined by a bonding function specifying the relationship between the amounts of the risky asset and a numéraire asset in a liquidity pool and the price of the risky asset on the centralized exchange follows an exogenous process that is assumed to be a semimartingale. Our first result shows there exists an equivalence between the bonding function of the decentralized exchange and its arbitrage profit function. We next rigorously define the notion of an arbitrage policy in our framework and its corresponding cumulative profit over a finite time horizon. We then study the problem of determining the optimal arbitrage policy to maximize the agent's expected cumulative profit over a finite time horizon. It turns out that the form of the optimal arbitrage policy varies depending on whether the price process is a martingale, sub- or super- martingale. These results provide insights into how blockchain latency depends on price dynamics. We also extend our results to settings including a discount rate as well as competition between agents.

3 - Genai and Corporate Disclosure: An Empirical Study**Yan Wu, Georgia Institute of Technology, Atlanta, GA, United States, Pan Li, Mingfeng Lin, Dongjun Wu**

The unprecedented capabilities of Generative Artificial Intelligence (GenAI) in text-generation, editing, and comprehension are driving its increasing adoption in corporate disclosures. In this study, we examine the relationship between GenAI and corporate disclosures using publicly available data. Specifically, we focus on the changes in the narrative style of corporate disclosures after Generative AI tools became available. Our preliminary results show changes in the sentiment of corporate disclosures. Practical implications for corporations, investors, as well as regulators are also discussed. This study provides insights into the shifting role of GenAI in corporate disclosures and its potential impact on market reactions.

4 - Governance Interlocks in Blockchain-Based Decentralized Autonomous Organizations

Jing Tian, The Pennsylvania State University, University Park, PA, United States, Ling Xue, Jingzhao An, Xi Zhao

A key phenomenon in the governance of decentralized autonomous organizations (DAOs) is member interlock, in which individual members concurrently engage in the governance of multiple DAOs. Although this practice resembles board interlock in traditional corporate governance, the understanding of individual-level interlock motivations and activities remains limited. In this study, we specifically consider the differential motivations of knowledge acquisition and knowledge utilization for DAO members' interlocks. We utilize a unique context of token airdrop used by Optimism Collective (OP) to attract members from other DAOs. Our analysis illustrates that members utilize their existing governance knowledge in other DAOs to OP governance, such as voting on related governance proposals. Furthermore, interlocking members acquire new knowledge from OP governance, which is applied subsequently in their home DAO governance. Leveraging the special setting to compare airdropped members and non-airdropped members, our analysis reveals that interlocking costs primarily hinder individuals' knowledge acquisition motivation for interlocking. Finally, we find that the interlocking DAO members can influence the voting outcomes of novel OP proposals, and token airdrop can foster this influence by facilitating the contributions from more interlocking members with knowledge acquisition motivation. Our study contributes to multiple streams of research on interlocks, DAO governance, and blockchain token airdrops.

TE10

Building A Level 3 A312

Digital Security and Trust

Contributed Session

Contributed

Chair: Farzaneh Moghadam, University of North Texas, 1418 W Oak Street, Denton, TX, 76201, United States

1 - Unveiling the Relationship Between Algorithmic Controls and Fairness Monitoring

Liuliu Qin, Old Dominion University, Norfolk, VA, United States, Yuming He, Lan Cao

As algorithmic management systems increasingly assume key managerial functions such as evaluation, direction, and discipline, concerns about fairness in the workplace have grown. This study examines the relationship between employees' perceptions of algorithmic control and their fairness monitoring behaviors. Drawing on the conceptual frameworks of algorithmic control and fairness monitoring, we investigate how distinct control mechanisms—recommending, restricting, recording, rating, replacing, and rewarding—influence employees' focused attention on distributive, procedural, and informational fairness. Through a two-study design involving scenario-based experiments and field surveys, our research contributes to the literature by integrating fairness monitoring into the algorithmic control discourse and identifying how distinct controls influence perceptions of justice. Practically, the findings can inform organizations on how to align control strategies with fairness expectations to foster trust and engagement.

2 - Monitoring Illegal Content on Digital Platforms

Jishnu Hazra, Indian Institute of Management Bangalore, Bengaluru, India, Tarun Jain

Illegal content uploads on social platforms have grown rapidly in recent years. We consider the situation in which along with the subscription-based legal content on the platform, there is free illegal content that may generate revenues through advertisements for the platform. We analyze three scenarios of illegal content monitoring. We study the optimal content pricing and illegal content monitoring strategies for the content developer and platform firm. We find conditions content monitoring strategies that generate higher payoffs for the digital platform and the content provider.

3 - Improper Modeling of Competing Risks: Impact on Risk Estimation and Algorithmic Fairness

Vincent Jeanselme, Columbia University, New York, NY, United States, Brian Tom, Jessica Barrett

Accurate time-to-event prediction is integral to decision-making in healthcare and beyond, influencing treatment recommendations, resource allocation, and policy guidelines. Survival analysis — the statistical framework used to model time-to-event data — accounts for patients who do not experience the event of interest during the study period, known as censored patients. However, many patients experience events that prevent the observation of the outcome of interest. These *competing risks* are often treated as censoring, a practice frequently overlooked due to a limited understanding of its consequences.

This work theoretically demonstrates why treating competing risks as censoring introduces substantial bias in survival estimates, leading to systematic overestimation of risk and potential disparities in clinical recommendations. First, we formalize the problem of misclassifying competing risks as censoring and quantify the resulting error in survival estimates. Specifically, we develop a framework to estimate this error and demonstrate the associated implications for predictive performance and algorithmic fairness. Further, we assess how differing risk profiles across demographic groups result in group-specific errors, potentially widening disparities in medical delivery. Our findings, supported by simulations and empirical analysis of cardiac event prediction, demonstrate that ignoring competing risks disproportionately impacts the most at-risk groups, potentially amplifying inequity.

By quantifying the error and highlighting the fairness implications of the common practice of considering competing risks as censoring, our work provides a critical insight into the development of time-to-event models: practitioners must account for competing risks to reduce disparities in risk assessment and downstream decision-making.

4 - Effects of Cybersecurity Readiness on Firm Performance: Evidence from Conference Calls

Minh Nguyen, Michigan State University, East Lansing, MI, United States

This work introduces a novel firm-level measure of cybersecurity, termed cybersecurity readiness, developed using text mining techniques. We then employ various empirical methods to estimate the causal effects of this cybersecurity readiness measure on firm performance. Our findings indicate that cybersecurity readiness positively influences firm performance, and this effect remains robust across multiple conditions. These results suggest that firms can benefit from investing in the cybersecurity aspects of information technology and data privacy.

5 - From Black Box to Boardroom: the Role of XAI in High-Stakes Decision-Making

Farzaneh Moghadam, University of North Texas, Denton, TX, United States

In today's fast-moving and uncertain business environment, leaders are increasingly turning to Artificial Intelligence (AI) to guide critical decisions. But as these systems grow more complex, they often become harder to understand—leaving executives to question not just what the AI recommends, but why. This study explores how Explainable AI (XAI) can help decision-makers feel more confident, informed, and in control.

We focus on two key human responses: how well leaders understand the AI's reasoning (AI Awareness), and how much they trust it. Drawing on Decision Theory, we propose that XAI improves decision quality not directly, but by increasing both awareness and trust. To test this idea, we designed an experiment where participants faced a simulated high-stakes decision using either a black-box AI system or one that provided clear explanations for its recommendations.

The results were striking: participants who saw how the AI reached its conclusions felt more assured and made better decisions—faster and with more clarity. Trust and awareness played a crucial role in this improvement.

Our findings offer more than just technical insight—they highlight the psychological bridge that explainability builds between humans and machines. For organizations investing in AI, the message is clear: empowering decision-makers means going beyond performance metrics. It means designing systems that people can understand, relate to, and rely on—especially when the pressure is high.

This research adds to the growing conversation around human-centered AI and offers a roadmap for making advanced technologies more effective, ethical, and usable in the real world.

TE11

Building A Level 3 A313

TSL Data-driven Research Challenge (Award Session 2)

Award Session

Transportation Science and Logistics (TSL)

Chair: Hai Wang, Singapore Management University, Singapore, Singapore

Co-Chair: Lei Zhao, Tsinghua University, Beijing, N/A

1 - Potential-Based Greedy Matching for Dynamic Delivery Pooling

Matias Romero, Columbia Business School, New York, NY, United States, Hongyao Ma, Will Ma

We study the pooling of multiple orders into a single trip, a strategy widely adopted by online delivery platforms. When an order has to be dispatched, the platform must determine which (if any) of the other available orders to pool it with, weighing the immediate efficiency gains against the uncertain, differential benefits of holding each order for future pooling opportunities. In this paper, we demonstrate the effectiveness of using the length of each job as its opportunity cost, via a potential-based greedy algorithm (PB). The algorithm is very simple, pooling each departing job with the available job that maximizes the savings in travel distance minus a half of its distance (i.e. the potential). On the theoretical front, we show that PB significantly improves upon a naive greedy algorithm in terms of worst-case performance: as the density of the market increases, the regret per job vanishes under PB but remains constant under naive greedy. In addition, we show that the potential approximates the marginal cost of dispatching each job in a stochastic setting with sufficient density. Moreover, we conduct extensive numerical experiments and show that despite its simplicity, PB consistently outperforms a number of benchmark algorithms, including (i) batching-based heuristics that are widely used in practice, and (ii) forecast-aware heuristics that estimate the marginal costs of dispatching different jobs using historical data.

2 - Analytics for on-Demand Food Delivery: A Hurwicz Perspective for Three-Sided Platforms

Qinghe Sun, The Hong Kong Polytechnic University, Hong Kong, Hong Kong, Ziao Zhao, Li Chen

Problem definition: We study operational decision-making in three-sided on-demand food delivery (OFD) platforms, which coordinate customers, restaurants, and couriers under uncertainty. The key challenge is to dynamically optimize order assignment and courier relocation to balance competing stakeholder objectives—minimizing delivery delays and expired orders while preserving courier availability and ensuring fair workload distribution—amid ambiguity in order arrivals and transit times. **Methodology/results:** We develop a multi-period, multi-objective optimization model that balances stakeholder-specific goals. To manage uncertainty flexibly and effectively, we propose a Hurwicz satisficing optimization (HSO) framework that determines the minimal level of optimism under which the decision-maker's objectives remain attainable despite ambiguity. Leveraging the nice properties of the proposed certainty equivalent, we derive tractable reformulations of the proposed HSO model. Implemented via a rolling-horizon mixed-integer program, the HSO model consistently outperforms heuristics and deterministic baselines in real-data experiments, achieving superior trade-offs among the performance dimensions. **Implications:** Our findings underscore the critical importance of accommodating uncertainties in balancing stakeholder objectives in a three-sided market. The results demonstrate the value of ambiguity-aware decision-making and offer actionable insights for platform operators. This work also contributes to the broader literature on robust and multi-stakeholder optimization in OFD operations under uncertainty.

3 - Optimization of Courier-Order Matching and Delivery Performance for in-TIME Delivery Systems

Keji Wei, Tongji University, Shanghai, China, People's Republic of, Yi Fan, Yufeng Cao, Xi Xiong

Meal delivery services face growing demand, making it essential for platforms to improve efficiency and delivery speed. To address this, we introduce the Adaptive Delivery Matching with Agent-based Courier Reinforcement Optimization (AD-MACRO) framework, which tackles the match-delivery decision problem for delivery platforms.

Our approach employs a two-layer decision framework designed for multi-agent environments, leveraging distributed reinforcement learning and optimization methods to manage decision-making challenges. To deal with the complexity of large numbers of agents, we propose the Grouped-Actor Policy Proximal Optimization (GAPPO) algorithm. GAPPO reduces decision-making complexity by grouping agents with similar characteristics, effectively handling the heterogeneity between platform-employed and crowdsourced couriers.

We validate the framework using real dispatch data from Meituan, showcasing its ability to enhance delivery performance. Additionally, to improve practical relevance, we incorporate couriers' rejection behaviors, modeled through fixed probabilities and economic incentives. By integrating these factors along with transportation equity considerations, our framework optimizes both order assignments and delivery performance. The results include a 77.61% reduction in delay rates and a 50.52% increase in couriers' leisure time.

This leads to an effective solution that benefits all parties involved—platforms, couriers, and customers—by improving efficiency, reducing delays, and ensuring fairer working conditions for couriers.

4 - Data-Driven Real-Time Online Platform Food Delivery Optimization: The Meituan Case

Nicolas Kuttruff, Technical University of Munich (TUM), Munich, Germany, GOLMAN RAHMANIFAR, MOSTAFA MOHAMMADI

Dynamic and uncertain environments in Online Food Delivery (OFD) systems present major challenges for real-time courier-order assignment, particularly under tight spatiotemporal constraints and behavioral uncertainties. Motivated by the 2024 INFORMS TSL–Meituan Data-Driven Research Challenge, this study proposes a real-time optimization framework for the courier-order assignment problem, formulated as a Continuous Markov Decision Process (CMDP). In contrast to batch-based dispatching, the CMDP supports event-driven, order-level decision-making, enabling agile responses to evolving system conditions. A Matching Degree (MD) function guides assignment by incorporating spatial feasibility, detour cost, workload balance, and timing penalties. To further enhance operational performance, a postponement mechanism defers assignments when advantageous, allowing the system to exploit future information and improve order consolidation. Courier acceptance behavior is modeled using a Discrete Choice Model (DCM), capturing context-specific acceptance probabilities based on delivery distance, time of day, and demand intensity. The CMDP framework is benchmarked against a Rolling Horizon Assignment (RHA) policy, which makes myopic assignments without considering future demand variations. Numerical experiments show that CMDP outperforms the benchmark RHA strategy with higher responsiveness (≤ 0.41 seconds), higher throughput (1,887 vs. 1,558 orders), and lower rejection rates (19.33% vs. 29.87%). The framework also scales efficiently, with minimal computational overhead as order volume doubles. Moreover, CMDP leads to higher courier utilization (78% vs. 15%), reflecting improved allocation effectiveness.

5 - Delay-Risk Minimizing Order Dispatch in Online Food Delivery Platforms

Ning Zhu, University of Science and Technology of China, Hefei, Israel

The rapid growth of the meal delivery industry and increasing market competition have intensified the need for efficient and reliable delivery operations to enhance customer satisfaction. This paper addresses the order assignment problem for on-demand meal delivery platforms, aiming to minimize delay risks under delivery time uncertainty. We propose a delay-risk minimization model based on two delay-related metrics: total expected delivery delay (TEDD) and joint delay probability (JDP), which capture the severity and frequency of late deliveries. To account for uncertainty, we adopt a distributionally robust optimization (DRO) framework with a mean-variance ambiguity set constructed from historical delivery data and predictions from a linear regression tree (LRT) model. The DRO model is reformulated into a linear assignment problem, enabling efficient real-time optimization and deeper analysis of delivery trade-offs. Numerical experiments on real-world data validate our model's effectiveness, showing that robust assignment strategies significantly reduce delay risks while balancing delivery efficiency and reliability. Unlike traditional delivery time-based objectives, which fail to translate shorter delivery times into lower delay risks, the delay-related metrics improve service reliability with only a minimal compromise in efficiency. These findings highlight the practical value of our approach for enhancing operational performance in dynamic delivery environments.

TE12

Building A Level 3 A314

Data-Driven Demand and Supply Planning

Contributed Session

Contributed

Chair: Hiruni Niwunhella, North Carolina State University, Raleigh, NC, United States

1 - Penalty Functions for Truck-and-Drone Routing Problems

Ke Xu, University of Southern California, Los Angeles, CA, United States, John Carlsson

Drones are increasingly being integrated into last-mile delivery networks to alleviate reliance on conventional ground trucks. This collaboration can involve various combinations of trucks and drones and incorporate diverse operational constraints tailored to specific delivery requirements. In previous studies, however, each problem variant is typically addressed by a specialized solution method. In this study, we introduce a versatile heuristic approach based on the Lin-Kernighan-Helsgaun algorithm with penalty functions to solve a wide range of drone-assisted delivery variants with a unified solution framework. Its flexibility and effectiveness are demonstrated through numerical experiments on benchmark instances designed for several representative problem variants, including the Flying Sidekick Traveling Salesman Problem (FSTSP), the Traveling Salesman Problem with Multiple Drones (TSP-MD) and the Vehicle Routing Problem with Drones (VRPD).

2 - Predict-Then-Optimize Inventory Placement in Modern Warehousing

Ahtesham Bakht, USF, Tampa, FL, United States, Hadi Gard

The evolution of modern warehousing—driven by e-commerce, robotics, and AI—signals a paradigm shift from static, rule-based inventory placement to dynamic, data-driven strategies. In this study, we propose a predict-then-optimize framework that integrates demand forecasting and pairwise co-order frequencies with integer programming to generate optimized inventory placement plans. Our approach enables smarter, more adaptive warehousing by aligning item locations with predicted order patterns. We demonstrate the effectiveness of our method through numerical experiments and a real-world case study focused on order picking efficiency.

3 - A Novel Approach for Dynamic Pallet Allocation in Next-Generation Logistics Systems

Dilhani Marasinghe, University of Tennessee at Chattanooga, Chattanooga, TN, United States, Pubudu Jayasekara

Efficient pallet assignment plays a vital role in optimizing warehouse operations within next-generation logistics systems. This study presents an optimization model for dynamic pallet allocation to both outbound trucks and storage locations, with the objective of minimizing total operational costs. The model operates within an open horizontal collaborative network, where facilities differ from traditional cross-docks and breakbulk terminals. Unlike conventional systems, freight in this setting lacks a predetermined route from origin to destination, requiring real-time, adaptive decision-making. Experimental results from a simulated environment representing this new type of logistics facility show notable improvements in efficiency and resource utilization over conventional rule-based strategies. To further enhance operational performance, we explore the use of a machine learning algorithm that enables the system to learn from historical patterns and adapt to dynamic logistics conditions.

4 - Optimizing Volunteer Deployment and Routing in Food Banking Systems with Equity Constraints

Hiruni Niwunhella, North Carolina State University, Raleigh, NC, United States, MehrSadat Salami, Leila Hajibabai, Faisal Alkaabneh

Efficient volunteer deployment and routing are critical for ensuring equitable and timely food distribution in non-profit food banking systems. This study presents an optimization approach to assign volunteers to service locations and design routing plans while upholding equity constraints. The objective is to minimize operational costs (efficiency) and ensure fair assignment distribution (equity), considering factors such as volunteer availability, service center demand, vehicle capacity, and time windows. The problem is defined as a vehicle routing problem with pickup and delivery with equity constraints (VRPPD-e) and formulated using a mixed-integer linear program. To address scalability and improve computational performance, heuristic refinements and decomposition strategies are incorporated to solve the problem. Specifically, we develop a customized branch-and-price algorithm incorporating Lagrangian relaxation to solve large-scale instances of the developed model. Numerical experiments conducted on a test network highlight the model's effectiveness in achieving both efficiency and equity. It is also extended to a large-scale implementation in Durham, NC, involving more volunteers, service centers, and operational complexities. This work offers a practical approach to optimizing volunteer logistics in food banking, supporting equitable and efficient service.

TE14

Building A Level 3 A316

EV Infrastructure: Planning and Optimization

Contributed Session

Contributed

Chair: Hamid Najafzad, One UTSA Circle, 78249

1 - Flood-Resilient Electric Vehicle Infrastructure: a Bi-Level Optimization Framework for Urban Transportation Networks

Wencheng Bao, University of Illinois Urbana-Champaign, Urbana, IL, United States, Eleftheria Kontou

Rapid growth in electric-vehicle travel, combined with more frequent flooding events, is challenging the resilience of urban-transportation-systems. The advent of electric road transportation and its infrastructure, like charging stations, underline the need for fortification against hazardous flooding events. We present a novel optimization framework that jointly prescribes road hardening, and charger retrofits to mitigate transportation systems performance reduction during flooding. The problem is set as a bi-level program: the upper level minimizes expected transportation system disruption costs subject to budget for retrofits, while the lower level models traveler response through a stochastic user equilibrium assignment that embeds: (i) flooding impact link travel times, (ii) M/G/K/∞ queue for en-route Level-2 and DC-fast-charging dwell times, and (iii) reachable-demand-chain that filters paths by battery range. We solve the lower level with Method-of-Successive-Averages and wrap the bi-level search in a tabu-guided-hyper-heuristic. The framework is implemented to the Chicago Metropolitan Area's road network. Workflow data include Sentinel-1-derived-flood-depth-rasters processed with the Floodwater-Depth-Estimation-Tool, K-means-clustered charger inventories, and Deep-Gravity-origin-destination matrix generation. Simulation over a 12-hours storm shows flooding event disables 41.4% of origin-destination pairs. Prescribed targeted repairs restore up to 94.3% of those trips while remaining within the retrofit budget. Incremental investments cut mean in-vehicle time from 153 to 77 mins and curb average charging queues from 33.9 to 8.44 mins. The analysis shows the value of behavior-driven prioritization for infrastructure fortification. The framework equips planners with a new model for allocating limited resilience funds across roads and charging assets as electric-vehicle adoption accelerates and hazardous events intensify.

2 - Scalable Optimization of EV Charging Networks: Combining Column Generation with RL-Driven Pricing

Iason Liagkas, University of Michigan Ann Arbor, Ann Arbor, MI, United States, Neda Masoud, Shiyue Zhao

The growing adoption of electric vehicles (EVs) necessitates a robust, strategically deployed charging infrastructure. This research introduces a two-level decision-making framework for optimal EV charger placement and dynamic pricing. We integrate mathematical optimization for charger siting with reinforcement learning (RL) for real-time price adjustments, addressing both long-term spatial allocation and short-term demand fluctuations. To solve the large-scale optimization problem, we employ a **column generation approach** that iteratively explores profitable charging configurations. Four specialized heuristics are designed to enhance this exploration: (1) **Greedy Demand-Based Heuristic**, prioritizing high-traffic zones based on historical demand; (2) **Dual Price-Guided Heuristic**, leveraging dual prices from the master problem to identify zones with the greatest marginal value; (3) **Hybrid Greedy + RL-Informed Sampling**, combining demand estimates with RL-driven profitability insights for refined zone selection; and (4) **Monte Carlo Sampling with Constraints**, exploring diverse configurations to discover non-obvious, high-potential placements. Each candidate configuration is evaluated using RL-based pricing strategies to determine profitability, ensuring only the most promising are added to the master problem. This joint optimization-pricing model provides a scalable, data-driven solution for EV infrastructure planning, dynamically responding to user behavior and competitive pricing. Our results demonstrate that this integrated framework maximizes revenue while ensuring accessibility, offering practical insights for urban mobility planners and investors.

3 - Optimizing Infrastructure and Charging Strategies for Electric Delivery Fleets Under Pv, Grid, and Cost Constraints

Joel Weingart, ZHAW - School of Engineering, Zurich, Switzerland, Maike Scherrer

The electrification of commercial delivery fleets introduces complex trade-offs between infrastructure investment, operational flexibility, and energy cost efficiency. This study presents a strategic optimization model designed to support long-term planning decisions for electric vehicle (EV) fleets and corresponding charging infrastructure. The model focuses on minimizing the Total Cost of Ownership (TCO) by jointly optimizing the sizing of photovoltaic (PV) systems, stationary battery storage, vehicle batteries, and grid connection capacity. It also incorporates fast and bidirectional charging capabilities to enhance system flexibility and reduce grid dependency.

Formulated as a Mixed-Integer Linear Programming (MILP) problem, the model evaluates a wide range of infrastructure configurations under realistic boundary conditions, using one-hour time steps. Key input data include high-resolution, year-long PV generation profiles, dynamic electricity tariffs, vehicle energy consumption patterns, and operational schedules derived from a real-world parcel logistics case study in Switzerland. The model is embedded in a scenario-based framework, allowing for the systematic exploration of infrastructure sizing options and their associated cost-performance trade-offs.

Results demonstrate how the temporal alignment between PV production and fleet availability, as well as the capacity to store and redistribute energy within the fleet, significantly influence cost optimization outcomes. The model highlights the value of bidirectional charging and seasonal planning in reducing peak grid demand and operational expenses. This work provides a scalable decision-support tool for logistics operators seeking to transition to electric fleets while maintaining cost control and energy resilience.

4 - Charging Station Operation Optimization for Middle-Mile Commercial EVs with Travel Behavior and Time-Varying Energy Costs

Yuta Tsutsumi, The University of Osaka, Suita, Japan, Katsuya Sakai, Hiroshi Morita

The adoption of electric vehicles (EVs) in logistics has been rapidly expanding, particularly in last-mile delivery operations. However, the transition to EVs for middle-mile deliveries, which refer to trips between distribution centers, remains limited due to the short driving range of EVs. To enable widespread use of EVs in middle-mile logistics, it is essential to consider not only the deployment of charging infrastructure but also its efficient operation. Although many existing studies have focused on the optimal placement of charging stations, research on optimizing their operation under dynamic electricity pricing remains scarce. In this study, we propose a mathematical optimization model to determine time-dependent charging prices at each station, aiming to minimize the total social cost, which includes vehicle travel distance, traffic congestion, and electricity consumption. The model assumes that each commercial EV independently chooses its route and charging schedule to minimize its own generalized cost, defined as the sum of travel time, travel distance, and charging price. By incorporating time-varying electricity prices that increasingly reflect fluctuations in renewable energy supply, lower charging prices can be interpreted as indicators of enhanced renewable energy utilization and reduced societal energy costs. To capture temporal variations in charging demand, we introduce time windows for charger availability. This addition enables the model to represent both the spatial and temporal dynamics of EV logistics operations. The proposed framework provides a socially optimal strategy for operating existing charging infrastructure and offers valuable insights into the integration of transportation and energy systems for sustainable logistics planning.

5 - A Two-Stage Stochastic Optimization Model for the Equitable Deployment of Fixed and Mobile Electric Vehicle Charging Stations

Hamid Najafzad, University of Texas at San Antonio, San Antonio, TX, United States, Omar Abbaas, Gabriel Nicolosi, Moddassir Khan Nayeem, Fuhad Ahmed Opu, Yara Altarawneh

A major barrier to wide Electric Vehicles (EVs) adoption is the absence of a reliable and equitable charging infrastructure. Rigid and poorly situated charging stations create coverage gaps and slow down EV adoption especially in underserved communities. This paper proposes a two-stage stochastic mixed-integer programming model for the optimal deployment of Fixed and Mobile Charging Stations (FCSs and MCSs) across multiple zones and periods. At the beginning, a finite dominating set of candidate locations is identified using the Edge Scanning Search (ESS) algorithm, an exact and continuous location method. We modify the ESS algorithm to incorporate existing public charging stations within the network, thereby avoiding redundant coverage and enhancing solution efficiency. In the first stage of our model, FCSs are allocated based on long-term traffic patterns, budgetary constraints, and socioeconomic factors to ensure stable baseline coverage. The second stage dynamically assigns MCSs in response to short-term demand fluctuations and uncertainties, aiming to minimize relocation costs while maximizing coverage. We use a scenario-based framework to capture variability in demand during peak and off-peak hours. Numerical experiments on realistic networks demonstrate the model's capacity to enhance system resilience and reduce unmet demand. Additionally, comparative results reveal that hybrid deployment strategies integrating both FCSs and MCSs outperform fixed-only approaches in terms of cost-efficiency, adaptability, and social equity. These findings offer practical insights for planners and policymakers focused on developing accessible and demand-responsive EV charging infrastructure.

TE15

Building A Level 4 A401

Ridesharing Operations

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Yifan Shen, University of Washington, SEATTLE, WA, United States

1 - Models to Balance Stability and Efficiency in Multi-Rider Ridesharing Systems

Hoda Atef-Yekta, James Madison University, Potomac, MD, United States, Mohsen Emadikhav

Most ridesharing research focuses on enhancing overall efficiency, such as reducing total travel distance or increasing platform profits. However, focusing only on efficiency does not ensure a stable solution. In shared rides, a rider's experience is shaped not only by cost but also by the combination of the co-riders, impacting the pickup time and the route distance. If a subset of dissatisfied riders can collectively improve their outcomes by deviating from the platform's assigned matches and form a new group, they create what is known as a blocking coalition. The presence of such coalitions can lead to instability, making the platform vulnerable to losing riders to competing services that offer more attractive ride options.

We propose two optimization paradigms to address stability in multi-rider ridesharing systems. In the first approach, we develop integer and mixed-integer programming models that explicitly encode stability requirements as constraints. In the second approach, we introduce a novel side-payment scheme, where monetary incentives are used to subsidize certain riders and discourage the formation of blocking coalitions. We highlight the effectiveness of our proposed approaches to balance efficiency and stability under various scenarios.

2 - On-Trip Matching for Shared Rides

Yifan Shen, University of Washington, SEATTLE, WA, United States, Julia Yan, Chiwei Yan

This research studies pricing and matching decisions for shared rides, where multiple passengers are pooled into the same vehicle. The predominant shared rides model is based on non-bipartite matching, and makes all matching decisions pre-trip: riders are either matched with each other before dispatch, or sent out alone, with no further matching allowed. This provides a parsimonious model of platform operations, but fails to explicitly consider the continued matching potential of riders initially dispatched solo – who may later be matched as compatible requests arrive.

We focus on on-trip matching, which allows riders to be dynamically matched with others while en route. In this setting, requests are dispatched immediately, avoiding rider waiting time, but have continued matching potential throughout their trip. This opens up the possibility of multi-party matching and enables the formation of trip chains, where a vehicle serves a sequence of pick-ups and drop-offs across compatible riders. While such strategies may be already heuristically used in practice, the dynamically varying matching values of riders on-trip are not captured in existing optimization models.

We aim to analyze the value of on-trip matching – identifying the conditions under which it improves system outcomes relative to pre-trip policies. We also study how platforms should price under this matching policy, and develop scalable algorithms to solve the pricing and matching problem.

3 - Two-Stage Matching in Ridehailing Systems

Xiyitao Zhu, University of Illinois Urbana-Champaign, Champaign, IL, United States, Hai Wang, Lavanya Marla

In nowadays ride-sourcing markets, it is common for ride-sourcing platforms (e.g., Uber, Lyft) to hold unserved passengers and idle vehicles for a short period of time and then batch match them for better systemic performances. The intuition is that better matching outcomes could be obtained in a denser bipartite graph with more supply and demand information accumulating over time through waiting. However, in current practice of the batch matching framework, such information accumulation is only captured by the batch matching time interval. To more effectively utilize information gained through waiting beyond the batch matching time interval, this work proposes a passenger-based two-stage matching framework, which features two stages of waiting with duration T_1 and T_2 enforced by the platform for each passenger. In this framework, each waiting passenger only receives an initial vehicle assignment from the platform after the first successful match and the platform could improve his final vehicle assignment later. This study models the system equilibrium of the dynamic matching process under the two-stage matching framework, and analyze how different system performance metrics (expected matching rate, expected waiting time, etc) are influenced by platform decisions, duration of the two stages and maximum matching radius. The theoretical findings are then validated by both analytical results, which are collected by solving a system of non-linear equations, as well as simulation results, which are collected from an agent-based simulator. It is shown that with proper selection of values for platform decisions, this two-stage matching framework improves performances of the ride-hailing systems.

4 - The Value of Learning in Peer-to-Peer Transportation Matching Problems

Ran Hu, Rensselaer Polytechnic Institute, Troy, NY, United States, James Bailey, Marlin Ulmer

We consider a dynamic matching problem where an online matching platform (e.g., UberEATS) assigns on-demand delivery requests to a fleet of crowdsourced suppliers/drivers. Throughout the day, customers arrive in the system and request delivery from stores to their home locations. Suppliers, operating independently, have the option to accept or decline delivery offers based on features such as location, distance, and expected compensation. However, online matching platforms do not have direct knowledge of the suppliers' choice/decision functions, which adds additional uncertainty to the online matching problem. To address this problem, we introduce online learning techniques to learn a supplier's choice function based on past accept/reject decisions to accurately predict supplier behavior. Motivated by methods in reinforcement learning, we create a closed feedback loop that balances speedy deliveries with preference learning, enabling adaptive and efficient supplier-customer matching under uncertainty.

5 - Who Gets to Ride? Comparing Transportation Eligibility Policies for Saanich Public Schools

Olivia Phillips, Georgia Tech, Atlanta, GA, United States, Arthur Delarue, Sebastien Martin

In public school transportation, eligibility policies determine which students get to ride the bus and which do not. Determining the operational cost of individual eligibility policies is an intractable problem and school districts must often make decisions with incomplete information. We collaborate with the Saanich School District in British Columbia, Canada, to develop methodologies to evaluate the costs and benefits of various eligibility policies. Our work involves both analyzing the current policy as well as designing optimization models that maximize utility while keeping costs below a given threshold.

TE16

Building A Level 4 A402

AI and LLM Applications for Intelligent Decision-Making

Invited Session

Analytics Society

Chair: Bahareh Kargar, New Jersey Institute of Technology, Martin Tuchman School of Management, Newark, NJ, United States

1 - Enhancing Meat Supply Chain Decision-Making Using Fine-Tuned Vision-Language Models and Robust Optimization

Bahareh Kargar, New Jersey Institute of Technology, Newark, NJ, United States, Majid Baqeri, Hamed Jahani

The meat supply chain (MSC) faces significant challenges due to the perishability of products, leading to substantial economic and environmental costs. Traditional monitoring and logistics methods often fail to provide real-time predictive insights needed to mitigate spoilage risks. In this study, we propose an AI-driven framework integrating fine-tuned vision-language models (VLMs) — specifically NVILA and Qwen-VL — with a robust optimization model for improving freshness prediction and logistics management in MSCs. Using the Meat Freshness Image Dataset, we fine-tune VLMs to classify meat into freshness categories and employ these predictions within a distributionally robust optimization (DRO) model to manage storage, transportation, and spoilage decisions under uncertainty. Our framework leverages Wasserstein-based DRO to model uncertainties in freshness, demand, and transportation times, leading to improved cost efficiency, service levels, and supply chain resilience. Extensive numerical experiments demonstrate that our approach significantly reduces worst-case costs and spoilage compared to nominal optimization methods, offering a scalable and effective solution for perishable supply chain management.

2 - Transforming Early Autism Detection by AI-Powered Video-Based Gait Analysis

Ziteng Wang, Northern Illinois University, Dekalb, IL, United States, Sinan Onal, Allison Gladfelter, Milijana Buac

Although children show clinical signs of autism early in their development, many children are not diagnosed until well into preschool years, and children from low-resourced, underserved, and rural backgrounds are diagnosed at an even later age. Moreover, current practices designed to detect autism rely solely on parent reports and clinician observations. There is a critical need for accessible, innovative, and more inclusive tools for early detection of autism. The objective of this research is to develop an easily accessible early autism screening web-based application to objectively assess children's walking gait to provide an autism index. The research is not intended to replace current diagnostic practices, but rather to offer a cost-effective, readily accessible way to screen for autism. We created a mobile-friendly website where a user (for instance, parent of the child being screened) can upload phone-recorded videos of the child walking, along with basic demographic information, and will receive an autism likelihood index, with an interpretation guide and suggested actions, as output. We use AI-powered gait models and deep learning models to predict the child's likelihood of autism. The preliminary test shows over 80% accuracy.

3 - Integrating Generative Adversarial Networks and Distributionally Robust Optimization for Green Manufacturing Capacity Planning

Xin Zhou, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Bo Zeng, Zhengsong Lu, Na Geng

Green manufacturing has become a strategic priority for firms seeking to enhance environmental sustainability and fulfill social responsibilities. This study addresses a real-world capacity planning problem in collaboration with a multi-factory manufacturer. We develop a two-stage distributionally robust optimization (DRO) model that considers both demand and climate-related uncertainties. To improve computational efficiency, a generative AI network is integrated into the solution scheme of DRO, leveraging AI's pattern recognition capabilities to enhance solving efficiency. The effectiveness of the proposed model and the AI-enhanced framework is validated through experimental analysis.

4 - Modeling Supply Chains via Graph Neural Networks

Zhaonan Qu, Columbia University, New York, NY, United States

5 - Debiasing Negotiation: Enhancing Intelligent Decision-Making and Human-AI Collaboration in Complex Supply Chain Deals

Vatsal Mitesh Tailor, University of Illinois Urbana-Champaign, Champaign, IL, United States, Shi Qiu

Negotiations in Supply Chain Management (SCM) are high-stakes, complex processes hindered by human cognitive limitations—such as anchoring and overconfidence—and the static nature of existing Decision Support Systems (DSS). This research proposes an AI-powered negotiation coach leveraging Large Language Models (LLMs) and reinforcement learning to augment human decision-making. The system analyzes real-time dialogues to interpret counterpart strategies, delivers context-aware recommendations, mitigates biases via feedback, and clarifies multi-issue trade-offs. Crucially, it also investigates embedded biases within AI negotiator and their impact on Human-AI collaboration for negotiation.

Supply chain contract negotiations are inherently uncertain and cognitively demanding, with human negotiators constrained by bounded rationality. While AI agents can enhance strategies and provide real-time support, their efficacy hinges on ethical model design and alignment with human preferences. Our study introduces a multi-agent LLM framework for SCM negotiations, evaluating its performance across diverse scenarios. Results demonstrate that human-AI collaboration improves efficiency and outcome alignment. By integrating dynamic behavioral analysis with bias-aware AI training, the framework addresses limitations of traditional DSS/NSS tools, which lack adaptive guidance. The work emphasizes proactive debiasing of LLMs—tackling inherited human biases and algorithmic distortions—to ensure equitable recommendations. It further explores optimizing Human-AI synergy, balancing AI-driven tactical insights with human judgment. This research advances intelligent negotiation tools by bridging behavioral science, AI ethics, and SCM, offering a pathway to mitigate cognitive barriers while safeguarding against AI-driven inequities.

TE18

Building A Level 4 A404

Military and Security II

Contributed Session

Contributed

Chair: Matthew Webb, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, 02139, United States

1 - AI and Cyber Powered Drone Warfare Innovations: Information-Theoretic Analysis of Emergent Cybersecurity Risks Across Military and Business Applications

Katya Drozdova, Seattle Pacific University, Seattle, WA, United States

Drone warfare is driving battlefield AI and business innovations. In particular, the Russia-Ukraine war has generated an unprecedented drone arms-race, including First-Person View (FPV) technologies as well as electronic warfare, offences and defenses across informational and physical domains. Existing commercial and military technologies have been adapted. Entirely new systems have emerged. These innovations are not only in technical design and practical deployment but also in intelligence, entrepreneurship, and production. For example, Ukraine's menacing Baba Yaga evolved from an agricultural drone; Russia's Upr 'Vampire' drone advanced from a volunteer enthusiast's hand-made creation through wartime testing to industrial production; 'Geranium-2' kamikaze drone is an iteration of Iranian 'Shahed' loitering munition build from a downed and reverse-engineered U.S. military Sentinel UAV. Innovations in AI-powered systems, naval drones, ground and underground robotics as well as civil-military collaborations have sprung up. Concurrently, an offense-defense arms race is revolutionizing electronic warfare, cybersecurity, evasion, and information operations. These developments are reshaping global conflicts and peacetime risks. Effective responses require strategic reorientation. In our cyberspace-reliant world—where public and defense IT infrastructures are interdependent—practically no one or nothing is safe from FPV drones and AI-powered systems. Moreover, cyberwarfare has outpaced civilian applications, threatening public safety and profitability of commercial drone markets, if risks are not properly addressed. Toward developing the necessary understanding, this research systematizes key technological advances, use cases, and vectors of emergent applications. Information-theoretic analysis is used to identify threat indicators and warning—with metrics for uncertainty reduction, risk management, and strategy optimization.

2 - Strengthening Intervention Design for Illicit Supply Chains: A Network Dynamics-Based Framework

Rashid Anzoon, University of Illinois Urbana-Champaign, Urbana, IL, United States, Rakesh Nagi, Chrysafis Vogiatzis

Illicit trade remains a persistent global threat, continuously seeking to expand through their covert supply chain networks. Efforts to disrupt these networks often yield limited success, as intervention in one area frequently triggers resurgence in another, a phenomenon commonly known as the balloon-and-cockroach effect. These unintended operational consequences complicate the design of effective intervention strategies, which must anticipate and counter the evolving nature of illicit supply chains. In this talk, we introduce a network dynamics-based framework that leverages structural vulnerabilities within the supply network to design targeted interventions. We also discuss optimal resource allocation strategies to maximize the impact of these efforts.

3 - Comparative Study of Calibration Frameworks for Flight Scheduling

Jared Erickson, OpsLab, Austin, TX, United States, Christopher Montez, Chris Biggins, Sujeevraja Sanjeevi

Effective flight scheduling for warfighters requires balancing multiple competing objectives, including mission fulfillment, equitable pilot workload, and timely pilot advancement. We present a comparative study of calibration frameworks designed to empower users to easily configure optimization models that reflect operational priorities of flying squadrons. The goal is to provide intuitive mechanisms for tuning objective coefficients, enabling users to obtain schedules that align with their expectations and squadron goals.

Our study evaluates multiple calibration approaches. These methods are assessed using historical scheduling data, with a focus on user interpretability, flexibility, and the quality of the resulting schedules. Rather than tuning coefficients automatically, our frameworks are designed to facilitate user-driven adjustment, supporting a more transparent and controllable scheduling process.

4 - Drone Routing in GPS-Denied Environments

Chase Murray, University at Buffalo, Amherst, NY, United States

While we have grown to rely on GPS for navigation, in many contexts GPS cannot be relied upon to provide localization. Several strategies have emerged for accurate drone navigation without the aid of satellites. Most of these involve some combination of computer vision processing fused with the measurements of the drone's onboard sensors (e.g., gyroscopes, altimeter, accelerometer, and compass). This talk defines a new problem where a drone must collect imagery data without the aid of satellite

navigation. The images collected during these runs may be subsequently used to provide more accurate navigation in future missions. Results from experiments with real drones validate the algorithm's performance.

5 - Agent-Based Modeling of Search Operations in Undersea Warfare

Nipun Galande, Virginia Tech , Blacksburg, VA, United States, Esra Toy, Anna Svirsko, Daphne Skipper

This work presents a novel agent-based simulation framework that integrates dynamic weather modeling with a Bayesian updating mechanism to enhance maritime search operations. We incorporate these Bayesian updates within the simulation and a Mixed-Integer Programming (MIP) model, comparing their solutions and performance under uncertainty. Our contributions include an innovative culling heuristic, pruning low-probability target paths, and extensive simulation replications to robustly quantify search performance under varying conditions. An Agent-Based Model (ABM) simulates and optimizes the dynamics of searchers and high-risk targets within a maritime environment. The ABM incorporates complex operational assumptions and constraints, facilitating a comprehensive analysis of search strategies and resource allocation. A key objective of this research is to address the impact of weather uncertainties on search operations, which a deterministic optimization model may not capture. Weather conditions, including wind speed, visibility, precipitation, and storm occurrences, are dynamically modeled to impact the operational capabilities of search assets. The simulation incorporates Areas of Uncertainty (AOUs) that expand over time after the detection of a submarine vessel by surveillance assets. These AOUs are also influenced by changing weather conditions, necessitating adaptive search strategies to maintain operational effectiveness. Our heuristic optimizes the search asset pathfinding process, enabling searchers to navigate AOUs more effectively by prioritizing areas with a higher probability of target presence, improved observation capabilities, and greater strategic value. The findings aim to inform the development of optimized search strategies and resource deployment plans, resulting in robust maritime security operations even in the face of environmental uncertainties.

6 - The Art of Hiding: Exploring Hiding and Rendezvous Tactics in a Contested Environment

John Becker, USNA, Annapolis, MD, United States, Samuel Zimmerman

Search is often studied as a symmetric game where the searcher's logic and payoff calculations inform how we expect the hider to behave. However, that is often not the case in practice as the hider's purpose and reasoning for behavior may not be in line with the searcher's reasoning. This research aims to explore and expand this facet of search which we call hiding theory. We look to answer how hiding policies can be evaluated and determine if optimal policies can be generated with limited information. Based on our findings, we hope to uncover general principles which are analogous to those found in search theory (e.g. "sweep rate"). Applications of hiding theory include how a downed pilot in enemy-controlled territory should behave to avoid detection and be rescued, how in-country terrorists may behave while awaiting orders, and how ballistic missile submarines should conduct deterrence patrols if they do not wish to be detected.

7 - Stop that Flow! Reinforcement Learning for Network Interdiction

Matthew Webb, Massachusetts Institute of Technology, Cambridge, MA, United States, Chris Caplice

Network interdiction problems are critical for modeling real-world security challenges including disruption of drug trafficking networks, military logistics during combat, and critical infrastructure protection. This presentation introduces a novel reinforcement learning (RL) approach to solve the maximum flow network interdiction problem for both deterministic and stochastic interdiction outcomes. By framing interdiction as a Markov decision process, our method trains RL agents capable of achieving comparable accuracy to mixed-integer optimization-based approaches, while achieving faster solve times for problems at scale. Our method also flexibly handles partial capacity reductions and multiple interdiction attempts without requiring structural reformulations of traditional optimization models. Unlike machine learning-assisted methods, which depend on pre-labeled training data and are vulnerable to labeling errors, this RL approach is immune to such issues as it learns interdiction policies through environment interaction rather than supervised data. This work bridges RL and combinatorial optimization, and offers a template for tackling other network interdiction problems beyond maximum flow.

TE19

Building A Level 4 A405

Learning, Control and Queues

Invited Session

Applied Probability Society

Chair: Xinyun Chen, Chinese University of Hong Kong, Shenzhen, Shenzhen, N/A

Co-Chair: Guiyu Hong, Shanghai University of Finance and Economics, Shanghai, N/A

1 - Online Learning for Dynamic Pricing with Delay-Sensitive Customers

Guiyu Hong, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Xinyun Chen, Yunan Liu

Motivated by sharing economy and online marketplaces, we propose an online learning approach to optimize a state-dependent pricing policy in service systems where customers are delay sensitive. Leveraging the underlying queueing structures, the proposed online learning method is featured with improved sample efficiency and faster convergence rate, compared to standard online learning algorithms.

2 - Exploring the Operational Potential of Queue-Ratio Policies in Multiclass Queueing Networks

Lucy Huo, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, Jim Dai, Jin Guang

Motivated by the recent stability analysis of multiclass queueing networks under queue-ratio policies by Zhao, Gurvich, and Hasenbein (2024), we build on their results to explore the operational implications of this policy class. Under their assumptions, together with a P -matrix reflection matrix, we establish that, in a multi-scale heavy traffic regime, the stationary distribution of the scaled workload converges to a product-form limit with exponentially distributed marginals. Together with state space collapse, the scaled queue lengths are also exponentially distributed in the limit. These results yield a tractable closed-form approximation for steady-state performance evaluation without simulation. Leveraging the closed-form approximations, we investigate their potential as a robust engineering tool for operational optimization, for example, identifying near-optimal policies for objectives such as delay minimization and resource balancing. We present numerical experiments and report our ongoing exploration in this direction, concluding with a discussion of open problems.

3 - Sample Path Large Deviations for Multivariate Heavy-Tailed Hawkes Processes

Xingyu Wang, University of Amsterdam, Amsterdam, Netherlands, Jose Blanchet, Roger Laeven, Bert Zwart

Hawkes processes capture the amplification and cascade effects of risks through temporal and spatial feedback, which prevail in various scientific, engineering, and business contexts. This talk develops sample path large deviations for multivariate Hawkes processes with power-law heavy tails in the mutual excitation rates. Our results address a broad class of rare events in Hawkes processes at the sample path level and, via the cluster representation of Hawkes processes, unravel the most likely configurations of (multiple) large clusters that could trigger the target events. These results also lay the foundation for precise theoretical insights and efficient rare-event simulation in practical systems subject to clustering and mutually exciting risks.

Our proof hinges on several technical developments. First, we provide a detailed probabilistic analysis of the spatiotemporal structure of large jumps in multi-type branching processes, which helps reveal the hidden regular variation that arises endogenously in the tail behavior of Hawkes process clusters. Next, we establish the asymptotic equivalence, in terms of M -convergence, between a suitably scaled multivariate Hawkes process and a coupled Lévy process with multivariate hidden regular variation. Along the way, we derive a sample-path large deviations principle for such Lévy processes, which is also of independent interest given the importance of Lévy models and the prevalence of hidden regular variation.

4 - Near-Optimal Control and Design of Critically Loaded Service Systems when Service and Patience TIMES are Dependent

Lun Yu, Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Ohad Perry

Customer behavior has significant impacts on the operations and performance of service systems. For example, even a small amount of abandonment, which is due to customer impatience, has dramatic impacts on key performance measures, and in turn, on the optimal staffing and control of such systems. However, in contrast to the prevalent models in the literature, the time that customers are willing to wait in queue for service often depends on their service requirement, and the impacts of this dependence can be sufficiently substantial so as to render those existing models futile.

In this paper, we consider large service systems that handle customers with dependent service and patience times, when those systems are staffed according to the Square-Root Staffing (SRS) rule in order to operate in the quality-and-efficiency (QED) asymptotic regime. We first develop bounds for the steady-state queue, delay and customer loss and show that the value of a single dependence scalar determines whether SRS guarantees that the system operates in the QED regime. We then propose design-and-control mechanisms that guarantee that SRS is near-optimal, in the sense that the QED regime is achieved. We further show that our near-optimality results are robust to the dependence structure, requiring no knowledge of the exact service-patience distribution, which is hard to estimate in practice. Finally, we provide simple methods for computing the optimal control parameters. The effectiveness of our analytical results and of our proposed solutions to the dependence-caused congestion are supported by extensive simulation studies.

TE20

Building A Level 4 A406

Bandits and Stochastic Control

Invited Session

Applied Probability Society

Chair: Ziv Scully, Cornell University, Ithaca, NY, United States

1 - Statistical Inference of Controlled Markov Chains: CLT for Transition Probabilities and Applications to Offline RL

Ziwei Su, Northwestern University, Evanston, IL, United States, Imon Banerjee, Diego Klabjan

We develop a central limit theorem (CLT) for the nonparametric estimator of the transition matrices in controlled Markov chains (CMCs) with finite state-action spaces. Our results establish precise conditions on the logging policy under which the estimator is asymptotically normal, and reveal settings in which no CLT can exist. We then build upon it to derive CLTs for the value, Q -, and advantage functions of any stationary stochastic policy, including the optimal policy recovered from the estimated model. Goodness-of-fit tests are derived as a corollary, which enable us to test whether the logged data is stochastic. These results provide new statistical tools for offline policy evaluation and optimal policy recovery, and enable hypothesis tests to distinguish whether observed data come from an MDP or a bandit.

2 - Artificial Replay and Thompson Sampling for Historical Data in Bandits

Christopher Martin, Yale University, New Haven, CT, United States, Siddhartha Banerjee, Sean Sinclair

We consider applications of the multi-armed bandit problem where some amount of historical data is available upfront, and additional data is collected adaptively through online interactions. Artificial Replay is a powerful meta-algorithm for bandit problems with historical data which has been shown to effectively reduce the use of redundant data without sacrificing regret when used with a wide class of underlying algorithms including UCB. In this paper, we apply the classic bandit algorithm Thompson Sampling to this historical-data bandit problem, and compare Artificial Replay to a full warm-start approach. We show a tight problem-dependent bound on the performance of Full-Start Thompson Sampling for K -armed Bernoulli bandits. We then complement this theoretical result with an empirical comparison of Artificial Replay and Thompson Sampling. In the K -armed bandit setting, while Artificial Replay does not guarantee the same regret as Full Start in general, it actually can outperform it in expectation under certain distributions of historical data. Additionally, we explore whether these findings apply to a broader context with empirical simulation of Gaussian Process Thompson Sampling and compare a Full-Start approach to our own Artificial Replay-style algorithm.

3 - Fooling Algorithms in Non Stationary Bandits

Gal Mendelson, North Carolina State University, Raleigh, NC, United States, Eyal Tadmor

This work is concerned with constructing worst case instances to prove lower bounds on the regret under various algorithms in non stationary bandit problems. We develop a new method for constructing such instances that fools algorithms to believe a sub optimal arm is optimal for a long period of time. This allows us to obtain new and better bounds than are currently known for popular algorithms such as UCB and sliding window UCB.

4 - Delayed Allocation in Marginalized Flow Models for Weakly Coupled Markov Decision Processes

Selva Nadarajah, University of Illinois Chicago, Chicago, IL, United States, Andre Cire

We introduce a process interpretation of fluid policies derived from Lagrangian relaxation of weakly coupled Markov decision processes (WDPs), revealing that existing approaches follow an "allocate-then-correct" sequence: first allocating marginal probability flows to component actions fluid linear program, then correcting for constraint violations. We identify that these feasibility corrections can be costly, degrading expected policy performance and increasing sample path variability, which we term fluid-model feasibility risk.

We propose delayed allocation, which reverses this process by identifying feasible actions before allocation. This creates a hierarchy of structure-aware fluid models that restrict actions to subsets of provably feasible solutions, operating on convex hulls of feasible action sets rather than linear relaxations. Our hierarchy converges in the limit to exact-feasible models from the literature, but remains computationally tractable at intermediate levels.

For a broad class of problems, we establish transferability results showing that approximation guarantees for deterministic constraint systems carry over to our stochastic formulations. We prove that polynomial-time approximation algorithms yield polynomial-time delayed allocation models, even when exact-feasible formulations require exponential representation – a striking complexity difference supporting intermediate models in our hierarchy.

Numerical experiments on dynamic assortment and maintenance problems demonstrate that delayed allocation reduces optimality gaps of standard approaches, while also decreasing performance variability. Our models also reduce reliance on frequent reoptimization, broadening applicability to environments where regular reoptimization is costly or impractical.

5 - Multiscale Replay: a Robust Algorithm for Stochastic Variational Inequalities with a Markovian Buffer

Milind Nakul, Georgia Institute of Technology, Atlanta, GA, United States, Tianjiao Li, Ashwin Pananjady

We introduce the Multiscale Experience Replay (MER) algorithm to solve stochastic variational inequalities (VIs) in settings where samples are collected in a Markovian buffer. MER utilizes a multi-scale scheme to emulate the behavior of independent and identically distributed (i.i.d.) samples, thereby mitigating the effects of data correlation without requiring knowledge of the Markov chain's mixing time. Our algorithm and its analysis demonstrate how the so-called "experience replay" heuristic from reinforcement learning can be applied in a provable and statistically efficient manner with dependent data. We discuss the implications of our results for temporal difference methods in policy evaluation and generalized linear models with dependent data.

TE22

Building A Level 4 A408

Stochastic Networks and Queueing Systems

Invited Session

Applied Probability Society

Chair: Guodong Pang, Rice University, Houston, TX, 77005, United States

1 - The Stationary Behavior of Reflecting Coupled Brownian Motions with Applications to Shortest Remaining Processing Time Queues

Amber Puha, California State University-San Marcos, San Marcos, CA, United States, Sixian Jin, Marvin Pena

With the objective to characterize the stationary behavior of the scaling limit for shortest remaining processing time (SRPT) queues with a heavy-tailed processing time distribution, as obtained in Banerjee, Budhiraja, and Puha (BBP, 2022), we study a collection of reflecting coupled Brownian motions (RCBM) indexed by the nonnegative real numbers. These RCBM arise by regulating certain coupled Brownian motions (CBM) to remain nonnegative. In the context of the BBP (2022) scaling limit, the regulated Brownian motion indexed by the nonnegative real number a represents the scaling limit of the amount of work in the system of size a or less. Thus, for the BBP (2022) scaling limit, the time t values of the RCBM describe the random distribution of the size of the remaining work in the system at time t . Our principal results characterize the stationary distribution of the RCBM in terms of a maximum process associated with the CBM. We obtain an explicit representation for its finite-dimensional distributions and a simple formula for its covariance. We further show that the RCBM converge in distribution to the maximum process as time approaches infinity. From this, we deduce the stationary behavior of the BBP (2022) scaling limit, including obtaining an integral expression for the stationary queue length in terms of the associated maximum process. While its distribution appears somewhat complex, we compute the mean and variance explicitly, and we connect with the work of Lin, Wierman, and Zwart (2011) to offer an illustration of Little's Law.

2 - Fluctuations of the Atlas Model from Inhomogeneous Stationary Profiles

Peter Rudzis, University of North Carolina, Chapel Hill, NC, United States, Sayan Banerjee, Amarjit Budhiraja

Brownian particle systems with singular (local time-mediated) interactions have important applications in queueing theory, mathematical finance, and models for physical surface growth. One such particle system is the Atlas model, a prototypical example in the class of rank-based diffusions from stochastic portfolio theory. Focusing on the infinite-dimensional version of this model, we establish fluctuation-type scaling limits for the occupation measure and order statistics of the system started from any stationary regime in which the particle density grows at an exponential rate. We show that the fluctuations can be described by an explicit Gaussian process, characterized as the solution to a stochastic partial differential equation (SPDE) in which the initial condition is Brownian motion, the linear operator is the infinitesimal generator of geometric Brownian motion, and the equation is driven by a certain additive space-time noise that is white in time and colored in space. The result yields precise insights about the large-scale properties of the system, including a simple and explicit description for the fluctuations of a tagged particle started from the bulk of the initial configuration.

3 - Some Large Deviation Principles for Load-Balancing Queueing Systems

Ruoyu Wu, Iowa State University, Ames, IA, United States, Amarjit Budhiraja, Eric Friedlander

We consider the large-scale load balancing queueing system under the Join-the-Shortest-Queue- $d(n)$ policy with diverging $d(n)$. The system consists of one dispatcher and n servers. When a task arrives at the dispatcher, $d(n)$ servers are chosen uniformly at random, and then the task is routed to the one with the shortest queue length. We establish large deviation principles for the system occupancy process, when $d(n)=n$ or in general $d(n)$ diverges. Proofs rely on certain variational representations for exponential functionals of Poisson random measures, stochastic controls, and weak convergence arguments. This is based on joint works with Amarjit Budhiraja and Eric Friedlander.

4 - Thompson Sampling Algorithm for Stochastic Games

Cohen Asaf, University of Michigan, Ann Arbor, MI, United States, Ruolan He, Yuqiong Wang

We study a stochastic differential game with N competitive players in a linear-quadratic framework with ergodic cost, where d -dimensional diffusion processes govern the state dynamics with an unknown common drift (matrix). Assuming a Gaussian prior on the drift, we use filtering techniques to update its posterior estimates. Based on these estimates, we propose a Thompson-sampling-based algorithm with dynamic episode lengths to approximate strategies. We show that the Bayesian regret for each player has an error bound of order $O(\sqrt{T \log(T)})$, where T is the time-horizon, independent of the number of players. This implies that average regret per unit time goes to zero. Finally, we prove that the algorithm results in a Nash equilibrium.

5 - Functional Limit Theorems for Pricing Dynamics in Limit Order Books with State-Dependent Multivariate Hawkes Processes

Ben Buran, Rice University, Houston, TX, United States, Bo Li, Guodong Pang

We study a tick-by-tick asset price model for a set of assets in one industrial sector (such as a technology sector) which is formed by a state-dependent Hawkes multivariate processes, representing the associated positive and negative jumps for each asset. In particular, both the baseline rate functions and the self-exciting functions in the stochastic intensities of the Hawkes processes depend on the price vector. We first show that the asset price model is well defined under certain regularity conditions. We then establish a functional law of large numbers (FLLN) and a functional central limit (FCLT) for the price dynamics. The FLLN limit for the price is given by the unique solution to a system of nonlinear ODEs. The FCLT limit for the price can be represented by the unique solution to a system of stochastic Volterra integral equations driven by Brownian motion, with the kernel functions depending on the FLLN limit for the price. The FCLT limit indicates both the path-dependent (memory) and state-dependent properties of the asset price dynamics. In the proofs, we develop a novel approach to establish the convergence of the LLN and CLT scaled processes, in order to tackle the challenges arising from the state dependency in the intensity processes.

TE23

Building A Level 4 A409

Advanced Analytics and Optimization Techniques

Contributed Session

Contributed

Chair: Ana Muriel, University of Massachusetts Amherst, Amherst, MA, United States

1 - Crossing the Rubicon of Rival Revelation: Dynamic Information Design in Competition

Yue Feng, Tianjin University, Tianjin, China, People's Republic of, Yunchuan Liu, Yi Liu, Jianxiong Zhang

In this paper, we study how firms strategically design information either about themselves or their rivals to influence consumer purchase decisions in dynamic competition. In particular, we focus on environments in which consumers' prior beliefs differ across the two firms' products, while consumers simultaneously face an outside option. We construct a dynamic Bayesian persuasion model in which a leader firm and a follower firm sequentially design and send their information to consumers to strategically shape consumers' beliefs on the two firms' products. In the model, both the leader firm and the follower firm can decide whether to manipulate consumers through information design, and whether to design to reveal information about either themselves or their competing rival. Our analysis shows that, strikingly, the leader firm may choose to "serve" the competing rival firm with a high risk gambler strategy of fully providing the rival information to consumers, especially when consumers have moderate prior beliefs for the leader firm. In addition, when consumers have relatively high prior beliefs for the leader firm, the leader firm may surprisingly choose a *downward twisted self-revelation* strategy, intentionally disseminating unfavorable signals about itself despite favorable product match, in a way to enhance the persuasiveness of its information. Finally, we show that an increase in competition may intensify information manipulation.

2 - Truthful Capacity Allocation Under Asymmetric Information: Leveraging Premium Pricing to Mitigate Ration Gaming

Burak Çelik, Aalto University School of Business, Espoo, Finland, Heikki Peura, FANG LIU

Ration gaming arises when downstream firms place orders exceeding their optimal quantity in response to a capacity shortage at their upstream supplier. Its negative effects can be mitigated through truthful allocation rules that incentivize firms to report their actual needs. We consider asymmetric downstream firms as they possess private information regarding their selling prices and operational costs, which determine their individual profitability. We propose novel capacity allocation mechanisms that utilize a premium wholesale price to provide priority allocation to downstream firms compared to those ordering at a standard wholesale price. We show that all the proposed mechanisms incentivize downstream firms to report truthfully, leading them to reveal their true demands in equilibrium. The first two proposed mechanisms prioritize fairness among the downstream firms with respect to allocation quantities, similar to uniform allocation. In contrast, the final mechanism overcomes a key limitation of existing mechanisms under private downstream information by maximizing supply chain profit in a decentralized supply chain, a performance level equivalent to a centralized system. This is accomplished by utilizing order quantities observed at premium wholesale price level to identify downstream firms with higher profit margins, and prioritizing capacity allocation to them. We subsequently characterize the strategies of the downstream firms and analyze the resulting equilibria for each mechanism. We evaluate the performance of the proposed mechanisms with respect to supply chain profit, using the uniform allocation rule as a benchmark. Our results also establish a connection between capacity allocation and auction theory via the Vickrey–Clarke–Groves mechanism.

3 - Strategically Robust Game Theory via Optimal Transport

Nicolas Lanzetti, ETH Zürich, Zürich, Switzerland, Sylvain Fricker, Saverio Bolognani, Florian Dörfler, D P

In many game-theoretic settings, agents are challenged with taking decisions against the uncertain behavior exhibited by others. Often, this uncertainty arises from multiple sources, e.g., incomplete information, limited computation, bounded rationality. While it may be possible to guide the agents' decisions by modeling each source, their joint presence makes this task particularly daunting. Toward this goal, it is natural for agents to seek protection against deviations around the emergent behavior itself, which is ultimately impacted by all the above sources of uncertainty. To do so, we propose that each agent takes decisions in face of the worst-case behavior contained in an ambiguity set of tunable size, centered at the emergent behavior implicitly defined. This gives rise to a novel equilibrium notion, which we call strategically robust equilibrium. Building on its definition we show that, when judiciously operationalized via optimal transport, strategically robust equilibria (i) interpolate between Nash and security strategies; (ii) come at no additional computational cost compared to Nash equilibria; (iii) often lead to better decisions and higher payoffs. Through a variety of experiments including bi-matrix games, congestion games, and Cournot competition, we show that strategic robustness protects against uncertainty in the opponents' behavior and, surprisingly, results in higher equilibrium payoffs – an effect we refer to as coordination via robustification.

4 - Timing the Life Cycle of Remanufactured Product Versions as New Product Generations Evolve

Ana Muriel, University of Massachusetts Amherst, Amherst, MA, United States, Yue Jin

As new product generations are launched, the firm must decide when to release the remanufactured version of the newest model and when to discontinue that of the previous one. We propose a steady state model that breaks the product generation interval into two stages. The new product is introduced at the beginning of Stage 1 and competes with the remanufactured version of the old generation. At some point, remanufacturing the new generation product will become feasible and the company will transition to Stage 2, offering remanufactured products of the new generation, which is more appealing to the customer. Two remanufacturing strategies are considered in the second stage: 1) offering only the new-generation remanufactured model, thus discontinuing the previous one, and 2) offering both the old and new remanufactured models. We quantify the benefits associated with the two-period strategy in steady state under the two options for old-gen discontinuation and analyze the timing decision for offering the new version of the remanufactured product, as well as the sensitivity to various model parameters.

TE24

Building A Level 4 A410

Information Operations in Social Media: Identifying and Countering Threats

Tutorial session

Tutorials

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

1 - Information Operations In Social Media: Identifying and Countering Threats

Tauhid Zaman, Yale University, New Haven, CT, United States, Yen-Shao Chen

The battlefield of information warfare has moved to online social networks, where influence campaigns operate at unprecedented speed and scale. As with any strategic domain, success requires understanding the terrain, modeling adversaries, and executing interventions. This tutorial introduces a formal optimization framework for social media information operations (IO), where the objective is to shape opinions through targeted actions. This framework is parameterized by quantities such as network structure, user opinions, and activity levels—all of which must be estimated or inferred from data. We discuss analytic tools that support this process, including centrality measures for identifying influential users, clustering algorithms for detecting community structure, and sentiment analysis for gauging public opinion. These tools either feed directly into the optimization pipeline or help analysts interpret the information environment. With the landscape mapped, we highlight threats such as coordinated bot networks, extremist recruitment, and viral misinformation. Countermeasures range from content-level interventions to mathematically optimized influence strategies. Finally, the emergence of generative AI transforms both offense and defense, democratizing persuasive capabilities while enabling scalable defenses. This shift calls for algorithmic innovation, policy reform, and ethical vigilance to protect the integrity of our digital public sphere.

TE27

Building B Level 2 B201

Strategic Decision-Making in Urban Systems: Optimization and Learning Approaches

Invited Session

Computing Society

Chair: Pavithra Sripathanallur Murali, George Mason University, Fairfax, VA, United States

1 - Uncertainty in Vulnerability of Metro Transit Networks: a Global Perspective

Nazanin Tajik, Mississippi State University, Mississippi State, MS, United States, Alireza Ermagun, Hani Mahmassani, Fatemeh Jannatabadi

This study measures the “uncertainty in vulnerability” of 50 metro transit networks in the most populated cities across the globe under benign and malicious attack scenarios. Uncertainty in vulnerability delineates the gap between the performance loss trajectory formed by link percolation under benign and malicious attacks. Three observations are discerned. First, vulnerability and uncertainty in vulnerability are a function of both size and physics of the network explained by connectivity measures. A 1% increase in the ratio of links to nodes increases the vulnerability by 0.50% and increases the uncertainty in vulnerability by 2.24%. A 1% increase in the ratio of the number of links to the maximum possible number of links decreases vulnerability by 0.03% and the uncertainty in vulnerability by 0.12%. Second, the topology of metro transit networks with <100 nodes follows one of the three analogous forms of tree-shaped networks, networks with one undirected cycle, or single depot networks, while the topology of metro transit networks with ≥100 nodes is closer to grid and matching pairs. Third,

metro transit networks (i) are less likely to resume the operation under malicious attacks, (ii) are more likely to resume the operation under benign attacks, and (iii) are susceptible to both severe and non-severe degradations under random attacks. Overall, it is shown that the most vulnerable transit networks experience the maximal uncertainty in vulnerability and own a topology analogous to a single depot.

2 - RL-Based Pricing for Decentralized EV Charging Control

Nima Rashgi Shishvan, University of Texas at Austin, Austin, TX, United States, Zhaomiao Guo

Electric Vehicles (EVs) offer a sustainable alternative to internal combustion engines by reducing greenhouse gas emissions. However, large-scale EV adoption poses challenges for power grid stability due to uncoordinated charging, which can lead to peak loads and accelerated infrastructure degradation. Centralized control approaches provide effective load management but are impractical due to privacy concerns. Pricing-based strategies formulated using bilevel optimization offer indirect control over charging behavior, yet requires full knowledge of users' travel patterns, which is often unrealistic. To address this limitation, we propose a Reinforcement Learning (RL)-based pricing framework that learns optimal charging incentives without any prior information about individual travel schedules. In this framework, the pricing agent adapts over time by interacting with EV users who make charging decisions based on past and current prices and personal constraints. Results show that, despite lacking access to user-specific data, the RL-based pricing policy achieves load distributions that closely approximate those of a full-information bilevel optimization benchmark. This demonstrates the potential of RL for scalable and privacy-preserving EV charging coordination.

3 - Simulation and Analytics for Safer, Greener, and More Equitable Transport Infrastructure

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

Transportation infrastructure is critical for urban mobility, yet its deterioration or failure poses complex and far-reaching challenges to resilience, sustainability, safety, and equity within large-scale networks. These challenges extend beyond immediate financial costs, impacting public health, daily accessibility, and the social well-being of communities. Simulation and analytics offer powerful, data-driven approaches to quantify, predict, and mitigate such impacts, especially where real-world data may be limited or incomplete.

This talk explores simulation-based methodologies for understanding and improving the performance of transportation infrastructure under both hazard and non-hazard scenarios. The hazard-focused case study models culvert (road-stream crossing) failures during flooding events. By simulating multiple failures across a representative county, we analyze their effects on network accessibility, route criticality, and the distribution of impacts across different neighborhoods, highlighting risks for populations that may already face mobility barriers. The non-hazard case addresses pavement deterioration and its effect on user costs, such as excessive fuel consumption, which can be a significant component of a network's environmental footprint. Here, simulation is used to reconstruct representative home-to-work commutes, generating actionable insights even in the absence of comprehensive travel surveys or individual trip records.

Both studies employ publicly available datasets and emphasize methodological transparency to ensure that results are robust and transferable to other urban contexts. Finally, the talk underscores the importance of accounting for equity in infrastructure planning and investment, demonstrating how simulation can reveal disparities and support more inclusive decision making for resilient, sustainable, and safe urban networks.

4 - How Long Should Supermarkets Wait to Donate? Optimal Policies for Donating Perishables with Profit and Food Waste Considerations

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

Food waste is a significant problem with economic, environmental, and social impacts. Supermarkets help reduce food waste by donating perishables nearing expiration, which reduces disposal costs, provides tax relief, and enhances their social image. Food banks re-purpose food that would otherwise be wasted by collecting, sorting, and distributing donations to food-insecure households through networks of charitable agencies. Conflicting objectives between supermarkets and food banks can create inefficiencies in the food bank supply chain that increase food waste. This paper examines supermarket donation policies and investigates the trade-off between supermarket profitability and total food waste in the food bank supply chain. We develop a stochastic dynamic programming model that specifies the optimal time and quantity of perishable items to donate under stochastic demand. We simulate a robust set of operating conditions to compare the optimal policy to two practical benchmarks, and evaluate the trade-off between profit and waste among the three policies. Our results show that the optimal policy not only maximizes supermarket profit but also significantly reduces total waste compared to last-minute donation strategies. Thus, higher supermarket profits and less waste in the supply chain can be achieved by the optimal policy, which shows that there is often no trade-off between two seemingly conflicting objectives. Lastly, it is especially important for policymakers to introduce high tax credits for items with lower holding costs to incentivize early donations and substantially reduce waste in the food bank supply chain.

5 - Optimizing Pricing Strategies Through Learning the Market Structure

Ezgi Karabulut, Sabanci University, İstanbul, Turkey, Khosroparviz Naderivarandi, Burak Gokgur

This study explores the integration of market structure learning into pricing strategies to maximize revenue in e-commerce and retail environments. We consider the problem of determining the revenue maximizing price of a single product in a market of heterogeneous consumers segmented by their product valuations; and analyze the pricing strategies for varying levels of prior market structure knowledge. In our problem, the seller sets the price of the product independently of the arriving customer profile; observes only the customer's decision to purchase or not depending on their respective product valuation; and updates the estimations for the unknown market parameters and the price for the new customer based on the purchase preference data. The proposed pricing mechanisms utilize customer preference data, and adopt a two-stage approach: an offline learning stage that focuses on learning market structure, including the number of segments and their sizes and valuations, and an online learning stage that prioritizes revenue maximization and iteratively updates these estimates to refine pricing strategies dynamically. In this seller-customer dynamic, we analyze the impact of different levels of prior market information on the collected revenue; and customize our algorithms for different information levels. Experimental results demonstrate the effectiveness of these methods in adapting to different levels of market knowledge, revealing the economic value of learning market structure in achieving near-optimal revenues. Our findings suggest that even with limited prior market knowledge, firms can benefit from incremental learning strategies to reach profit levels that are relatively close to those achieved in full-information scenarios.

TE28

Building B Level 2 B202

Software for Optimization Modeling and Deployment 2

Invited Session

Computing Society

Chair: Susanne Heipcke, FICO, Xpress Optimization, Birmingham, United Kingdom

Co-Chair: Robert Fourer, AMPL Optimization Inc., Evanston, IL, United States

1 - Accelerating Optimization Workflows Across All Platforms with Fico Xpress Workbench and Insight Docker Images

Alexander Biele, FICO, Darmstadt, Germany

Deploying optimization software often involves challenges like software dependencies and compatibility issues. FICO Xpress Workbench and Insight Docker images address these needs, enabling seamless deployment across platforms. This eliminates installation hassles, enhances user experience, and allows developers to focus on optimization.

This talk will highlight the benefits of Docker, demonstrating the ease of setup and minimal configuration needed for a unified development experience across Workbench and Insight. We will showcase the end-to-end deployment of a Python-based optimization model into a fully operational application with Xpress Insight, including remote

debugging capabilities for scenarios on a remote server.

This presentation will provide practical insights into leveraging Docker for optimization workflows, making development and deployment more efficient. Additionally, we will share key lessons learned from distributing pre-configured Docker images tailored for client use and training purposes.

2 - Automated Discovery of Optimization Algorithms

Ilias Mitrai, The University of Texas at Austin, Austin, TX, United States, Nikolaos Sahinidis

We propose a novel mathematical programming-based formulation for the automated discovery of optimization algorithms. Specifically, we model iterative algorithms as an expression tree of fixed depth. This modeling approach can represent many functional forms for the update function, i.e., a superstructure of candidate algorithms using simple mathematical operators (addition, multiplication, division, subtraction). This formulation allows the search for new update functions that satisfy a given convergence criterion while a desired objective, such as the complexity of the update function, is optimized. The resulting problem is a Mixed Integer Nonlinear Programming problem, which can be solved with state-of-the-art global optimization solvers such as BARON. We use the proposed approach to discover new first-order optimization algorithms for unconstrained optimization problems.

3 - A Cpu/Gpu Implementation of Randomized Rounding for Symbolic Regression

Xingjian Tao, Georgia Institute of Technology, Atlanta, GA, United States, Nikolaos Sahinidis

Symbolic regression aims to uncover interpretable, closed-form expressions that map input variables to observed outputs. While state-of-the-art solvers—such as Operon and PySR—rely on genetic programming, recent work has approached symbolic regression as a mixed-integer nonlinear program (MINLP) and introduced STAR, a randomized rounding algorithm to obtain high quality solutions of the MINLP. In this work, we develop a novel GPU-accelerated implementation of STAR. We describe the algorithms utilized, including redundancy elimination, constant optimization, and regularization. Finally, we present extensive computational results with problems from SRBench.

4 - Mathematical Programs with Embedded Surrogate Models Using Gamspy

Adam Christensen, GAMS Development Corp, Fairfax, VA, United States, Steven Dirkse

Recent advances in ML/AI have commoditized the development of surrogate models using tools such as PyTorch, Scikit-Learn, and TensorFlow. These surrogate models simplify inherently non-linear phenomena, approximating complex behaviors so they can serve as constraints in optimization frameworks. Embedding these models in algebraic modeling languages (AMLs) like GAMS remains challenging: designed for sparse algebra, AMLs lack seamless integration with third-party software. The rise of Python in data science has motivated a paradigm shift, inspiring tools that bridge classical AMLs and current computational techniques. We introduce GAMSPy, a native Python AML combining the mathematical transparency and scalability of traditional AMLs with Python's ecosystem. Its set-driven constructs and operator overloading preserve the syntax of handwritten algebra while supporting dense matrix operations—matrix multiplication, transposition, norms—essential to ML/AI. While the GAMS “classic” engine excels at indexed algebra, GAMSPy extends its capabilities to accommodate ML workflows. We demonstrate embedding a neural network trained in PyTorch to model an energy system as a constraint within an optimization problem, enabling system engineers to optimize plant operations with detailed energy conversion models. This workflow exemplifies applications spanning weather forecasting and market behavior modeling. We also compare GAMSPy to existing approaches, discuss future developments, and highlight innovative intersections of mathematical modeling and machine learning. GAMSPy represents a significant convergence of AML rigor and Python-driven ML versatility. Its design prioritizes computational efficiency, syntactic clarity, and scalability, offering a robust platform that overcomes integration hurdles and unlocks new possibilities at the intersection of optimization and data science.

5 - Towards Intelligent, Adaptive, and Explainable Strategy Designing Through Neurosymbolic AI: Integrating LLMs and Symbolic Reasoning

Mayank Bhalerao, University of Oklahoma, Norman, OK, United States, Siavash Mandegari, Janet Allen, Farrokh Mistree

Modern supply chains are increasingly exposed to disruption from geopolitical shocks, cyber threats, and climate-related events. However, most current strategy tools remain static, optimization-driven, and opaque; unable to adapt or explain decisions in dynamic conditions. This creates a critical gap in supply chain resilience, where strategic choices must balance adaptability with traceability and managerial alignment. In this paper, we propose a conceptual neurosymbolic framework for strategy design, combining Large Language Models (LLMs) for disruption perception, Graph Neural Networks (GNNs) for risk propagation, and symbolic planners for constraint-based decision generation. We model strategy as an evolving Cyber-Physical-Social System (CPSS), where physical assets, digital intelligence, and human governance co-evolve. We validate the framework through a detailed, hypothetical disruption scenario in an electronics supply chain, demonstrating how the system senses, reasons, plans, and explains in real time. Through our approach we enable satisfying-based, explainable strategies that adapt to uncertainty while aligning with operational and managerial goals. In this paper we address the following research question - How can strategy be designed, modeled and supported as an evolving Cyber-Physical-Social System through a neuro-symbolic AI framework that balances adaptability, traceability, and cognitive boundedness under disruption?

Keywords: Neorymbolic AI, Large Language Models, Evolving Cyber Physical Social System, Supply Chain, Strategy Design

TE29

Building B Level 2 B203

Industrial Loads as Flexible Resources in the Future Power Grid

Invited Session

ENRE: Energy

Chair: Constance Crozier, Georgia Institute of Technology, Atlanta, GA, United States

Co-Chair: Elnaz Kabir, Texas A&M University, College Station, TX, United States

1 - Understanding Resource Variability in High Renewables Systems to Inform Design of Flexible Industrial Technologies

Micah Ziegler, Georgia Institute of Technology, Atlanta, GA, United States, Aliza Khurram, Shannon Huang, Jessica Trancik

In energy systems that rely primarily on solar and wind power, resource fluctuations can both challenge efforts to meet demand reliably and potentially present opportunities for flexible technologies. Thus, there is considerable interest in deploying technologies that can modulate their use of electrical energy. In this research, we characterize the dynamics of fluctuations in solar and wind resources to understand how industrial processes could be designed to increase utilization of electricity infrastructure and ease a transition to high renewables systems. Using four decades of high-resolution weather data, empirical estimates of power demand, and a parsimonious energy system model, we simulate the operation of cost-minimizing, high renewables systems that include energy storage. From these simulations, we identify common characteristics of times when potentially excess energy might be generated, accounting for high uncertainty regarding technology costs and the overall level of reliance on renewable resources and storage. We further quantify how these characteristics might change under different scenarios, which include deploying additional energy storage, expanding transmission infrastructure, and relying on alternative power generators. The results provide insights that can guide the design and operation of industrial processes to enable greater utilization of renewable energy resources and inform efforts to plan and build energy systems to help mitigate climate change.

2 - On the Potential of Electrified Supply Chains to Provide Long Duration Demand Flexibility

Rina Davila Severiano, Georgia Institute of Technology, Atlanta, GA, United States, Constance Crozier

Demand flexibility can offset some of the variability introduced on the supply-side by variable renewable generation. However, most efforts (e.g. control of residential vehicle charging) focus on short durations -- typically on the scale of minutes to hours. This paper investigates whether a fully electrified supply chain (transport and manufacturing) could provide demand flexibility over longer durations, exploiting the latency that typically exists between the processing of raw material to the delivery of finished product. Using a case study of the cement industry along the East Coast of the United States, we demonstrate that electrified supply chains could shift gigawatt-hours (GWh) of electricity demand for durations of more than a week, largely following wind power variability. Furthermore, we show that this occurs using low levels of carbon taxing (below \$50/t), at which battery storage is not economically viable. A sensitivity analysis shows potential to provide flexibility in all considered cost

scenarios, although where the flexibility comes from can change (e.g. transport vs manufacturing). We show that today's cost of electrified heavy goods vehicles are the most significant parameter -- with substantially lower costs yielding a more demand-flexible supply chain.

3 - Towards Energy-Resilient Pharmaceutical Supply Chains: Integrating Network Design with Power Systems Modelling

Miriam Sarkis, Texas A&M Energy Institute, College Station, TX, United States, Halil Iseri, Eleftherios Iakovou, Efstratios Pistikopoulos

As power grids transition towards renewable integration and decentralization, industrial demand-side flexibility becomes a key enabler of grid stability. In the pharmaceutical industry space, manufacturing steps are often energy-intensive and many efforts in decarbonizing pharmaceutical supply chain operations have increased the reliance on renewable energy grids. This effectively increases the potential exposure of manufacturing networks to renewable energy variability and pressures decision-makers to develop mitigation strategies against uncertainty, including flexible adjustments to production schedules. Renewable energy-related variability is combined with inherent process and demand uncertainty for novel pharmaceuticals, highlighting a critical need for systematic frameworks towards energy-resilient and uncertainty-aware operation for the decarbonization and digitalization of the industry.

To this end, we present a framework based on mixed-integer linear programming and a rolling horizon solution methodology for pharmaceutical supply chain design and planning integrating power systems modelling. Given technology costs and capacity constraints, pharmaceutical demand profiles and parameter uncertainties, the optimization problem is solved to investigate the flexibility of supply chain networks in meeting demand targets, minimizing cost and leveraging on available energy and material-related production and storage capacity. The impact of uncertainty is also considered through a scenario-based approach, enabling the investigation of the potential bottlenecks emerging in pandemic scenarios.

Results illustrate a critical trade-off with respect to cost, demand satisfaction and reliance on the renewables energy grid. We highlight the significance of this systematic trade-off approach for informed decision-making in establishing increasingly electrified and digitalized pharmaceutical supply chains.

4 - Supply Chain Design and Manufacturing Planning Exploring Spatio-Temporal Variability of Electricity Prices

Elnaz Kabir, Texas A&M University, College Station, TX, United States

Current supply chain designs assume continuous energy availability, making them vulnerable to energy market volatility and disruptions. These risks are further amplified by the growing electricity demand from data centers, AI, and the resurgence of domestic manufacturing. At the same time, the global transition to variable renewable energy (VRE) sources introduces significant spatiotemporal variability in energy supply, deepening reliance on weather conditions and grid infrastructure. This evolving landscape demands a new generation of supply chains that are resilient, adaptable, and tightly coupled with dynamic energy systems. In this work, we develop a unified, data-driven framework to optimize energy-intensive supply chains by integrating energy use directly into manufacturing and logistics decisions. Our approach explicitly accounts for spatiotemporal variability in energy supply over extended time scales (e.g., weeks to months), which is often overlooked in existing studies. The framework supports dynamic production scheduling and storage optimization. Using real-world case studies, we demonstrate how our model improves operational resilience and energy adaptability in supply chain networks. Results show that by co-optimizing logistics and energy-aware manufacturing decisions, firms can significantly reduce costs, mitigate energy-driven disruptions, and better align operations with variable energy availability. This research provides actionable insights for industries seeking to future-proof their supply chains in the face of increasing energy uncertainty.

TE30

Building B Level 2 B204

Carbon Emissions and Green Innovation

Contributed Session

Contributed

Chair: Aqi Dong, Embry-Riddle Aeronautical University, Daytona Beach, FL, United States

1 - Radical vs. Incremental Green Innovation: Differential Effects on Environmental Misconduct

Prem Thapamagar, University of Science and Technology of China, Hefei, China, People's Republic of, Yan Zhou, Nisar Ahmad, Noor ul Amin, Ali Akbar, Mohammad Eqbal Rizaie

Drawing upon the Natural Resource Based View of a firm and the Fraud Triangle theory, we develop a framework for examining the effects of incremental and radical green innovations on environmental misconduct. Analysing a panel dataset on Chinese firms, we find that radical green innovation significantly reduces the likelihood of environmental misconduct, whereas incremental green innovation increases it. This relationship operates through two distinct mechanisms: (1) both innovation types enhance ESG performance, though radical innovations show stronger effects, which deter firms from committing environmental misconduct; and (2) radical innovation reduces financial constraints, which leads to a decline in the likelihood of environmental misconduct; incremental innovation increases them, creating conditions that the Fraud Triangle suggests may motivate misconduct. These findings demonstrate how organizational ambidexterity in innovation strategy affects corporate environmental behaviour. For managers and policymakers, our results suggest that prioritizing radical green innovations may simultaneously advance environmental compliance, legitimacy, and financial stability, resolving the apparent paradox between economic and ecological objectives more effectively than incremental approaches alone.

2 - Smart Contract and Carbon Adoption Levels in Supply Chains: a Two-Echelon Model

JINHO CHA, Gwinnett Technical College, Georgia, GA, United States

The integration of environmental sustainability practices and technological innovations such as smart contracts is reshaping modern supply chains. This paper explores the interaction between smart contract adoption and carbon adoption levels within a two-echelon supply chain framework, focusing on the retailer's decision-making process with multiple suppliers. We introduce a dynamic pricing model for suppliers, where prices are influenced by market conditions, demand fluctuations, and smart contract adoption, alongside fixed and variable costs. We also account for production capacity constraints and the flexibility of suppliers to scale operations based on demand and technological advancements. Given the inherent uncertainties in supply chain operations, including demand variability and risk factors, we propose a model incorporating stochastic demand processes (Pareto, Weibull, Lognormal, and Exponential distributions) and risk-adjusted procurement costs and production capacity. The model optimizes procurement decisions for the retailer, balancing the costs of carbon adoption and smart contract implementation with the need to meet customer demand and maintain efficiency. This framework provides valuable insights into how supply chain managers can make better decisions under uncertainty, incorporating both cost reductions from smart contracts and the challenges of environmental sustainability. Our findings highlight the trade-offs between operational efficiency and sustainability goals, offering actionable recommendations for retailers integrating smart contracts and carbon adoption strategies.

3 - Carbon Fee Policy with Credit Guarantee Mechanism in Taiwan

Cheng-Feng Wu, National Taipei University, Taipei, Taiwan

In the global trend towards net-zero emissions, businesses must simultaneously respond to government carbon policy regulations while facing supply chain issues. To achieve net-zero goals and align with international standards, Taiwan has pioneered a carbon fee policy as a carbon pricing mechanism. However, for countries with numerous small and medium-sized enterprise suppliers (such as Taiwan), this not only involves regulatory and cost considerations arising from the policy but also requires careful thought on selecting appropriate green suppliers to ensure sustainable development of businesses and supply chains. This study focuses on the buyer-manufacturer's perspective, and examine how buyer-manufacturers can use incentive contracts to reveal suppliers' private information and employ credit guarantee mechanisms to assist financially constrained suppliers and reduce supply risks, thereby achieving optimal financing and procurement decisions under the carbon fee policy. Specifically, this study constructs a financing and procurement model that considers the carbon fee policy's impact while addressing issues of supplier information asymmetry, financial constraints, and supply risks. In the model, buyer-manufacturers design incentive contracts to identify suppliers' efficiency attributes, namely production costs and carbon emission efficiency. Meanwhile, for financially constrained suppliers, buyer-manufacturers can better share risks with banks through credit

guarantee mechanisms after identifying suppliers, thus help them obtain necessary operating capital and achieving a triple win situation. Through model analysis, this study explores how policy parameters, macroeconomic factors, and supply chain elements influence buyer-manufacturer decisions, providing theoretical foundations for Taiwan's carbon fee policy design and practical recommendations for government and businesses during implementation.

4 - Collaborative Transportation Under Uncertainty with Emission Considerations

Secil Savasaneril, Orta Dogu Teknik Universitesi, Ankara, Turkey, Banu Yüksel

We examine coalition formation among shippers who co-load to optimize cost savings. Using a cooperative game model under uncertainty, we analyze the stability of the cost game through core existence, proposing allocation schemes. An emission game is defined to address the problem of a fair allocation of emissions while minimizing costs. Factors that ensure stable allocations of cost-savings and emissions are identified and the performance of the allocation schemes in attaining the core are assessed.

5 - Characterizing Farmer Heterogeneity in Smallholder Carbon Sequestration Programs: a Hierarchical Bayesian Approach

Aqi Dong, Embry-Riddle Aeronautical University, Daytona Beach, FL, United States, Shanan Gibson, James Gibson

Smallholder carbon sequestration programs represent a rapidly growing segment of global carbon markets, yet the heterogeneity among participating farmers poses significant challenges for portfolio risk assessment and program optimization. Using comprehensive data from 11,000 farmers in The International Small Group and Tree Planting Program (TIST) across Kenya, Uganda, Tanzania, and India, we develop a hierarchical Bayesian framework to identify farmer typologies and quantify uncertainty in carbon credit delivery. The dataset encompasses 102 variables spanning socioeconomic indicators, agricultural practices, carbon metrics, and community characteristics. We employ Bayesian mixture modeling to uncover latent farmer classes, revealing distinct profiles based on income (ranging from \$0.50 to \$15.80 USD per day), land holdings, crop diversity, and tree planting patterns. Hierarchical models capture uncertainty at multiple scales—individual farmers nested within communities, communities within districts—while accounting for measurement error in tree counts and growth rates.

Our analysis addresses a critical gap in carbon market operations: the lack of rigorous methods for assessing heterogeneity-driven risk in smallholder programs. By characterizing farmer typologies and associated uncertainty distributions, we enable portfolio construction strategies that optimize the risk-return profile of carbon credit investments. The framework has immediate applications for corporate procurement decisions, where understanding delivery variance is essential for meeting emissions reduction commitments. Beyond carbon markets, the methodology provides insights into the intersection of climate mitigation and rural development, demonstrating how statistical rigor can enhance both environmental and social outcomes in sustainability initiatives.

TE33

Building B Level 2 B207

Frontiers in Healthcare Analytics

Invited Session

Health Applications Society

Chair: Hossein Hejazian, Harvard University, Cambridge, MA, 02138, United States

1 - Causal Dag-Informed Reward Design

Ziping Xu, Harvard University, Allston, MA, United States

In reinforcement learning (RL) problems, different reward functions may lead to the same optimal policy, while some reward functions can be substantially easier to learn. In this talk, we propose a framework for reward design by constructing surrogate rewards with mediators informed by causal directed acyclic graphs (DAGs), which are often available in real-world applications through domain knowledge. We show that under the surrogacy assumption, the proposed reward is unbiased and has lower variance. Specifically, we propose an online reward design agent that adaptively learns the target surrogate reward in an unknown environment. Feeding the surrogate rewards to standard online learning oracles, we show that the regret bound can be improved. The improvement is evaluated through simulation studies based on mobile health datasets, where RL is used for digital interventions delivery.

2 - Multi-Location, Dynamic Staff Planning for a Healthcare System

Kumar Rajaram, University of California-Los Angeles, Los Angeles, CA, United States

We develop and implement a multi-stage robust optimization model to dynamically assign anesthesiologists across hospital locations under uncertain demand. The model incorporates on-call flexibility, fairness, and location constraints, and is solved using a nested column-and-constraint generation method. Implemented at the University of Pittsburgh Medical Center, it achieves 12% annual cost savings (~\$800,000) and demonstrates the value of forecast accuracy and staffing flexibility. The framework is generalizable to other healthcare workforce planning problems, such as nurse staffing.

3 - Does Order of Diagnosis Matter? Data-Driven Insights into Comorbid Hypertension and Diabetes Treatment Planning

Ezgi Gursel, University of Tennessee, Knoxville, TN, United States, Faezeh Bagheri, Supriya Sarker, Nafiseh Payani, G. Anthony Wilson, Bing Yao, Anahita Khojandi

This study analyzes treatment trajectories in patients with comorbid hypertension (HTN) and type 2 diabetes mellitus (T2DM) using retrospective cohort data from multiple Tennessee health systems. It aims to characterize differences in treatment planning as a function of the order in which the conditions are diagnosed. By exploring how treatment planning may be influenced by diagnosis order, this study seeks to identify patterns in clinical decision-making. Insights from this work may inform more personalized and effective care planning for patients managing comorbid conditions.

4 - A Numerical Framework for Describing the Transient Dynamics of Mt/Mt/1 Queues

Lingfei Zhang, Texas A&M University, College Station, TX, United States, Hrayr Aprahamian, Hadi El-Amine

This talk examines the transient behavior of queueing systems with non-stationary Markovian processes. We propose a novel, scalable Fourier-based approach to numerically characterize transient behavior under any functional form of the arrival and service processes. Our approach provides both the full distribution of the number of entities in the system and the expected sojourn time. Through extensive numerical experiments, we demonstrate that our technique accurately captures the transient behavior of non-stationary queues with high precision. Furthermore, we embed our approach within an optimization framework by introducing a Bayesian optimization procedure that efficiently optimizes stochastic systems with a limited number of function evaluations. Finally, we present a case study on college counseling centers using real data, showing that our approach identifies effective server allocation strategies, reducing total customer sojourn time by nearly 60% compared to current practices.

TE34

Building B Level 2 B208

Machine Learning and Optimization

Contributed Session

Contributed

Chair: Yeonjun Choi, Pohang University of Science and Technology, Seoul, Korea, Republic of

1 - Per-Group Distributionally Robust Optimization (Per-Gdro) with Learnable Ambiguity Set Sizes via Bilevel Optimization

Jangho Park, Sungkyunkwan University, Suwon, Korea, Republic of, Seobeom Jung, Woojae Lee, Sophia Hamm, Jihun Hamm

Group structures frequently influence model behavior; however, group membership is often unobserved during inference, limiting explicit control over group-specific performance. This can result in models performing well on certain groups but underperforming in others, leading to concerns about fairness. Moreover, each group may

follow a different distribution, and a subset of groups may be more susceptible to distributional shifts due to external factors, such as policy changes or environmental variation. To address these challenges, we propose a Per-Group Distributionally Robust Optimization (Per-GDRO) framework. It ensures fairness across groups and robustness to group-specific distributional shifts. In this framework, a ϕ -divergence ambiguity set governs adversarial group reweighting, and Wasserstein ambiguity sets capture local uncertainty within each group. We then develop an iterative algorithm that alternates between model updates and adversarial distributions across and within groups. Worst-case distributions over ambiguity sets can be directly determined through convex optimization. We employ a derivative-free surrogate optimization method to determine the size of these ambiguity sets in an adaptive manner. The experimental results demonstrate that the proposed method consistently enhances worst-group accuracy while preserving intra-group robustness.

2 - Machine Learning-Powered Lagrangian Relaxation for a Location-Partition Problem

Wen Luo, Huazhong University of Science and Technology, Wuhan, China, People's Republic of, Shiming Deng, Chu Chu

Instant retailers are increasingly establishing micro-fulfillment centers (MCs) to effectively serve their designated nearby communities. This location-partition problem can be formulated as an integer programming model and solved using Lagrangian relaxation (LR). To enhance the computational efficiency of repeatedly optimizing the relaxed sub-problem in LR, we propose a novel method that integrates ML techniques within the LR framework. Our ML-powered LR method maintains the LR framework while adopting pre-trained ML models to predict the most challenging aspects of the relaxed sub-problem. By utilizing feasible solutions derived from these predictions, this integration leverages surrogate subgradient theory to find a proper surrogate subgradient direction. In contrast to conventional end-to-end ML strategies, our approach significantly reduces reliance on the training and prediction accuracies of ML models and simplifies the collection of training samples by imitating a smaller-scale relaxed sub-problem. We explore various ML techniques, including decision trees and neural networks, to validate the proposed method. Experimental results indicate that the ML-powered LR method can achieve time savings of up to 48.62%, while also potentially yielding superior outcomes compared to the benchmark LR.

3 - Machine Learning and Regularization-Based Outer Approximation for Convex Minlp and Steel Production Optimization

Shengnan Zhao, National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang, China, People's Republic of, Lixin Tang, Fengyuan Shi

In this study, we introduce a novel outer approximation algorithm enhanced by machine learning and regularization techniques, designed specifically for convex Mixed Integer Nonlinear Programming (MINLP) problems. This approach employs machine learning to analyze cut generation patterns during iterations of the outer approximation algorithm, integrating the regularization information of the master problems feasible solutions to significantly enhance convergence efficiency. We test the effectiveness of the proposed algorithm on a plate design problem which is derived from real world application, as well as a set of models from an MINLP benchmark library. The numerical results show advantages of using machine learning and regularization combined with the outer approximation algorithm.

4 - Learning Acceleration Algorithms for Fast Parametric Convex Optimization with Worst-Case Guarantees

Rajiv Sambharya, University of Pennsylvania, Philadelphia, PA, United States, Jinho Bok, Nikolai Matni, George Pappas

We introduce a machine-learning framework to learn hyperparameter sequences for accelerated first-order methods (*e.g.*, the step size and momentum sequences in accelerated gradient descent) to quickly solve parametric convex optimization problems while maintaining worst-case performance guarantees. Robustness with respect to the problem parameter is incorporated directly into the learning process by augmenting the training objective with a regularization term that certifies worst-case performance over a broad class of problems (*e.g.*, the minimization of all L -smooth convex functions). The regularization term is derived from the Performance Estimation Problem (PEP) framework and relies on solving a semidefinite program (SDP) in which the algorithm hyperparameters appear as problem data. We show how to differentiate through this regularization term in the computational architecture, enabling the use of gradient-based training that balances the performance over the parametric family and worst-case guarantees over a much larger set of parameters. We showcase the effectiveness of this learned acceleration framework on a variety of numerical examples that span several different first-order methods including the accelerated versions of gradient descent, proximal gradient descent, and the alternating direction method of multipliers. Applications include parametric problems from signal processing, control, and statistics. Notably, our approach is highly data-efficient in that we only use ten training instances in all of the numerical examples.

5 - Predicting Patterns of Forecasting Failures in AI-Enabled Devices

Matthew Bobea, Institute of Service Science, National Tsing Hua University, Hsinchu, Taiwan, Galit Shmueli, Soumya Ray

Forecasting is used in a vast array of decision-making systems, including those based on AI-enabled devices such as wearable health monitors, smart energy grids, and smart home systems. The heavy reliance on time-series forecasting in these systems has given rise to an urgent need to address extreme forecasting failures that significantly impact organizations. Such errors can arise due to the nature of forecasting algorithms and their interaction with data patterns. In response to this challenge, we introduce TreeAlert, a novel framework designed to identify, interpret, and forecast patterns of extreme forecasting errors. TreeAlert uses regression trees to detect failure-prone conditions and explicitly links them to contextual and timing patterns to help improve models. We translate these conditions into human-readable rules to generate actionable insights and facilitate stakeholder understanding. TreeAlert forecasts future failure periods, thereby providing a forward-looking tool that empowers decision-makers to proactively anticipate vulnerabilities. We demonstrate TreeAlert on heart rate data from an IoT device. By bridging error analysis with forecasting we offer an interpretable solution for enhancing forecast driven support systems and for minimizing risks.

6 - Training Neural Networks with User-Defined Dimension-Specific Unimodality

HYUNHO KIM, KAIST, Daejeon, Korea, Republic of, Jong-Seok Lee

Unimodal functions, characterized by a single extremum, facilitate global optimization and play a crucial role in various mathematical fields such as optimization theory and probability. In this study, we propose a novel neural network architecture for effectively approximating unimodal functions over a user-specified subset of features. To estimate these dimension-specific unimodal functions, we decompose a network into monotonic and convex components and then recombine them to approximate a quasi-convex function for each constrained dimension. The monotonic and convex properties are achieved through specially designed units that utilize exponentiated weights and suitable activation functions. These units are integrated into a partially connected neural network structure, which ensures unimodality in the designated dimensions without compromising the network's expressive capability. Numerical experiments illustrate the effectiveness and practical applicability of the proposed approach.

7 - Bilevel Container Relocation Problem

Yeonjun Choi, Pohang University of Science and Technology, Pohang, Korea, Republic of, Insoo Park, Seokyun Hong, Kangbok Lee

We study a practical yet underexplored extension of the container relocation problem (CRP) where a limited number of pre-marshalling relocations are allowed before retrieval begins. This anticipatory step enables proactive improvements to downstream operations. We formulate the problem as a bilevel optimization model: the upper level selects pre-marshalling moves, while the lower level executes a given CRP policy—optimal or heuristic—on the resulting configuration. To our knowledge, this is the first work to model and solve such problems using a unified bilevel framework.

We develop an efficient exact tree search algorithm that explores all pre-marshalling sequences up to a given depth. We incorporate pruning and dominance rules based on structural properties to reduce computation without sacrificing optimality. The framework supports arbitrary CRP solvers, ensuring both generality and practical relevance. Computational results show that dominance rules improve efficiency, and our method outperforms MIP-based approaches under comparable time limits. The proposed framework offers a unified and efficient approach to container yard planning with realistic operational flexibility.

TE35

Building B Level 2 B209

Access and Operations in Public Infrastructure

Contributed Session

Contributed

Chair: Jorge Acuna, Universidad Adolfo Ibáñez, Viña del Mar, Chile

1 - Using Bin Packing Optimization to Investigate the Transportation Challenges Impeding School Start Time Changes in Wake County Public Schools

Warren Kay, United States Military Academy, West Point, NY, United States, Brandon McConnell, Kristin Thoney-Barletta, Russell King, Russell Nelson

In 2014 the American Academy of Pediatrics (AAP) recommended later school start times (after 8:30 AM) for adolescents to promote their health and academic success. Wake County Public School System (WCPSS) is the largest school district in North Carolina, with nearly 200 schools and over 160,000 students, and all its middle school and high school students start school at 8:30 AM or earlier. In the 2023–2024 school year, a national bus driver shortage exacerbated daily student transportation challenges and delayed the school board's ability to prioritize school start time changes. This study uses a mathematical optimization model following a bin packing approach to minimize the number of buses required to complete all the routes in the nine sub-districts within Wake County. Scenarios that changed the school start times to abide by AAP recommendations were generated and tested using bus route information from the 2023–2024 school year. Results show that areas with a higher concentration of magnet and special schools require a greater increase in the number of buses needed to swap elementary and high school start times. The results also indicate significant interactions between the sub-districts, which challenge WCPSS's ability to logistically isolate a sub-district to use as a case study for start time changes. This research investigates the current relationship between transportation requirements and school start times. It provides analysis that could inform decision-making and planning as WCPSS hires more bus drivers and pursues the best outcomes for its students.

2 - A PPP-Based Planning Model for Enhancing Customer Experience at A State-Managed Transport Facility

N Ravichandran, Professor (Retired), Indian Institute of Management Ahmedabad, Ahmedabad, India, Nikunj Kumar Patel

This study presents a strategic planning model developed for the Gujarat State Road Transport Corporation (GSRTC) to optimize bus port infrastructure through Public-Private Partnerships (PPP), aiming to enhance customer experience in state-managed transport facilities. Drawing insights from 15 PPP projects executed in two phases, the research formulates a simulation-based decision model that allows sensitivity analysis across key parameters such as land allocation, terminal operations area, and commercial space utilization.

The model distinguishes between two core development components: (i) the Bus Terminal Facility (BTF), leased for 30 years, and (ii) the commercial complex, leased for 90 years. A detailed financial framework estimates construction costs, operational expenses (maintenance, security, utilities), and potential revenue streams (commercial leases, advertising, parking). The Net Present Value (NPV) and profitability for concessionaires are calculated to inform bidding and policy decisions. In parallel, GSRTC's potential revenue realization and strategic land use planning are integrated within the model's outputs.

Key reflections include the need for GSRTC to preserve long-term operational flexibility, adopt shared service arrangements for housekeeping and security, and secure revenue shares from commercial operations. The model transforms transport terminals into multifunctional public spaces—enhancing passenger amenities and creating urban value through retail, entertainment, and social infrastructure.

This integrated modeling approach offers a scalable blueprint for transport authorities to structure PPPs in a financially viable, socially beneficial, and operationally sustainable manner. It balances public objectives and private incentives, turning transport nodes into catalysts for community-centric urban regeneration.

3 - Workforce Scheduling for Efficient and Equitable Access to Education

Mohsen Emadikhiaiv, Florida Atlantic University, Boca Raton, FL, United States, Carlos Cardonha, Mahyar Eftekhari, David Bergman

Nonprofits supporting schools with extracurricular programs face unique challenges in efficiently assigning their skilled trainers. This research stems from our collaboration with New York Sun Works (NYSW), a nonprofit organization committed to delivering sustainability science and climate education to K-12 students across New York City. In this talk, we present optimization models and algorithms designed to address variants of the workforce scheduling problems encountered by NYSW.

4 - A Bi-Objective Robust Optimization Model for Equitable Allocation of Field Hospitals in Surgical Backlog Reduction

Jorge Acuna, Universidad Adolfo Ibáñez, Viña del Mar, Chile, Sofía Zúñiga

In 2023, over 330,000 patients in Chile awaited elective surgeries in the public healthcare system, with wait times frequently exceeding 400 days. This situation emphasized persistent disparities in surgical access and the need for capacity expansion strategies. To address this challenge, we formulated a bi-objective robust optimization model for the strategic allocation of field hospitals and the coordination of interregional patient transfers. The proposed model simultaneously maximizes the number of patients treated (efficiency) and minimizes regional disparities in access to care (equity). Disparities were quantified using a Gini index focused on unmet surgical needs, highlighting functional differences across health regions rather than population size. Demand uncertainty was addressed through a budgeted uncertainty set, ensuring solution feasibility under adverse conditions. Employing the triangle splitting method, we constructed the Pareto frontier of non-dominated solutions. We applied the Nash Bargaining Solution to guide policy selection, treating the two objectives as cooperative players negotiating a Pareto-optimal compromise. Using real healthcare data from Chile, we evaluated various investment levels and examined how model-driven transfer decisions influenced system performance. Results show that combining field hospitals with targeted patient transfers leads to improved regional balance and system throughput. This framework provides robust, data-driven insights to support mid-term planning for surgical backlog reduction and more equitable healthcare delivery.

TE36

Building B Level 2 B210

Recent Advances in Stochastic Nonlinear Optimization II

Invited Session

OPT: Nonlinear Optimization

Chair: Sen Na, Georgia Institute of Technology, Atlanta, 30339, United States

Co-Chair: Miaolan Xie, University of Chicago, Chicago, IL, United States

1 - Reliable and Adaptive Stochastic Optimization in the Face of Messy Data

Miaolan Xie, University of Chicago, Chicago, IL, United States

Solving real-world stochastic optimization problems presents two key challenges: the messiness of real-world data, which can be noisy, biased, or corrupted due to factors like outliers, distribution shifts, and even adversarial attacks; and the laborious, time-intensive requirement of manually tuning step sizes in many existing algorithms. I will introduce a simple adaptive optimization framework that avoids the need for manual step size tuning by adaptively adjusting the step size in each iteration based on the progress of the algorithm. To address the issue of messy data, the algorithm only assumes access to function-related information through probabilistic oracles, which may be biased and corrupted. This framework is very general, encompassing a wide range of algorithms for both unconstrained and constrained optimization. It is applicable to multiple problem settings, such as expected loss minimization in machine learning, simulation optimization, and derivative-free optimization. Under reasonable conditions on the oracles, we provide an analytical framework that bounds the iteration complexity with high probability for settings with highly noisy and corrupted gradients.

2 - Towards Parameter Free Decentralized Optimization

Gesualdo Scutari, Purdue University, West- Lafayette, IN, United States

We consider decentralized optimization where multiple agents minimize the average of their (strongly) convex, smooth losses over a communication graph. Convergence of the existing decentralized methods generally hinges on an apriori, proper selection of the stepsize. Choosing this value is notoriously delicate: (i) it demands global knowledge from all the agents of the graph's connectivity and every local smoothness/strong-convexity constants—information they rarely have; (ii) even with perfect information, the worst-case tuning forces an overly small stepsize, slowing convergence in practice; and (iii) large-scale trial-and-error tuning is prohibitive. This work

introduces a decentralized algorithm that is fully adaptive in the choice of the agents' stepsizes, without any global information and using only neighbor-to-neighbor communications--agents need not even know whether the problem is strongly convex. The algorithm retains strong guarantees: it converges at linear rate when the losses are strongly convex and at sublinear rate otherwise, matching the best-known rates of parameter-dependent methods.

3 - Online Statistical Inference of Constrained Stochastic Optimization via Sequential Quadratic Programming

Sen Na, Georgia Institute of Technology, Atlanta, GA, United States

Constrained stochastic nonlinear optimization problems have attracted significant attention due to their ability to model complex real-world scenarios in physics, economics, and biology. As datasets continue to grow, online inference methods have become essential for enabling real-time decision-making without storing historical data. In this talk, we introduce an online inference procedure for constrained stochastic optimization by leveraging a method called Sketched Stochastic Sequential Quadratic Programming (SSQP). As a direct generalization of sketched Newton methods, SSQP approximates the objective with a quadratic model and the constraints with a linear model at each step, then applies a sketching solver to inexactly solve the resulting subproblem. Building on this framework, we propose a new online inference procedure called random scaling. Specifically, we construct a test statistic based on SSQP iterates whose limiting distribution is free of unknown parameters. Compared to existing online inference procedures, our approach offers two key advantages: (i) it enables the construction of asymptotically valid confidence intervals, and (ii) it is matrix-free, meaning that the computation involves only primal-dual SSQP iterates $(\mathbf{x}_t, \lambda_t)$ without requiring any matrix inversions. We validate our theory through numerical experiments on nonlinearly constrained regression problems and demonstrate the superior performance of our random scaling method over existing inference procedures.

4 - Neural Collapse and AI Optimization

X.Y. Han, Chicago Booth, Chicago, IL, United States

In the performance-dominated landscape of AI development, systematic understanding is difficult: Any architecture is fair game as long as it can climb the leaderboard. Yet, even among the overwhelming multimodality of neural networks, one motif remains: data is transformed layer-by-layer and iteration-by-iteration into high-dimensional representations with which predictions are eventually made. Thus, examining the geometry of these representations offers a powerful lens for the analysis of AI. Neural Collapse is a striking example, evident late in the training of predictive AI: Originally observed in classification networks during a 'Terminal Phase of Training' (TPT) – where training error vanishes but loss continues decreasing – Neural Collapse reveals a fundamental inductive bias. It encompasses four interconnected geometric regularities in the final layer representations: (NC1) Within-class variability of activations collapses towards zero as they converge to their class means; (NC2) These class means arrange themselves into a maximally separated, symmetric structure (a Simplex Equiangular Tight Frame); (NC3) The classifier vectors align with these class means in a self-dual configuration; (NC4) The model's decision mechanism effectively simplifies to Nearest Class-Center rule. This emergent geometric simplicity is linked to improved generalization, robustness, and interpretability. I will discuss the core Neural Collapse phenomenon, its theoretical underpinnings, and recent updates and extensions since its discovery.

5 - Online Scaled Gradient Methods

Madeleine Udell, Stanford University, Stanford, CA, United States

We introduce a framework to accelerate the convergence of gradient-based methods with online learning. The framework learns to scale the gradient at each iteration through an online learning algorithm and provably accelerates gradient-based methods asymptotically. In contrast with previous literature, where convergence is established based on worst-case analysis, our framework provides a strong convergence guarantee with respect to the optimal stepsize for the iteration trajectory. We obtain the first convergence guarantee for the widely-used hypergradient descent method (HDM) and for new variants with heavy-ball momentum (HDM-HB) and Nesterov momentum. HDM-HB significantly outperforms other adaptive first-order methods, and often matches the performance of L-BFGS using less memory and cheaper iterations.

TE37

Building B Level 2 B211

Recent Advance in Optimization Under Uncertainty: Theory and Applications

Invited Session

OPT: Optimization Under Uncertainty

Chair: Karmel Shehadeh, University of Southern California, Los Angeles, CA, United States

1 - Multistage Stochastic Programming for Hurricane Relief Logistics Planning with Rolling Forecast Uncertainty

Yongjia Song, Clemson University, Clemson, SC, United States, Sudhan Bhattarai

In this talk, we will discuss multi-stage stochastic programming (MSP) models and solution approaches for humanitarian relief logistics planning in hurricane disasters. Specifically, we study how the rolling forecast information can be integrated in an MSP model via the Martingale Model of Forecast Evolution (MMFE) to provide optimal adaptive logistics decision policies. We investigate state-space modeling to address the high-dimensional state space as a result of the MMFE model. Our numerical results and sensitivity analyses demonstrate the value of MSP for hurricane relief logistics planning compared to approaches that utilize rolling forecast information in a rolling-horizon fashion. We also investigate various trade-offs between policy flexibility, solution quality, and computational effort, from which we gain valuable managerial insights for the practitioners.

2 - Capacitated Spatial Newsvendor

Xian Yu, The Ohio State University, COLUMBUS, OH, United States, Jinhui Han, Ming Hu

We revisit the classical newsvendor problem by introducing capacitated supply and spatial transportation in a setting where a firm needs to position available supplies across a service region to meet random demand over the space. Leveraging optimal transport theory, we derive optimality conditions that minimize the sum of aggregate newsvendor loss and transportation expenses in a general framework, accounting for demand over continuous space or discrete locations or a combination of both. The solution structure retains the simplicity of the classical newsvendor solution but also differs due to the interplay of demand, transportation, and supply constraints. The demand effect drives supply placement to high-demand areas, the transportation effect prioritizes nearby demand, and the supply constraint restricts the ability to fully meet all demand, especially when the total supply is insufficient. The solution is explicitly characterized using Laguerre cell division in a specific semi-discrete setting, where initial supplies are stored at discrete facilities and demands are distributed over a continuous space. This structure reveals how transportation costs adjust the optimal demand quantile and how limited local supply influences this quantile through Laguerre cells' weight parameters. The derived optimality conditions further inform the development of numerical algorithms in both semi-discrete and fully discrete settings.

3 - Sequential Quadratic Optimization for Solving Expectation Equality Constrained Stochastic Optimization Problems

Haoming Shen, University of Arkansas, Fayetteville, AR, United States, Yang Zeng, Baoyu Zhou

A sequential quadratic programming method is designed for solving general smooth nonlinear stochastic optimization problems subject to expectation equality constraints. We consider the setting where the objective and constraint function values, as well as their derivatives, are not directly available. The algorithm applies an adaptive step size policy and only relies on objective gradient estimates, constraint function estimates, and constraint derivative estimates to update iterates. Both asymptotic and non-asymptotic convergence properties of the algorithm are analyzed. Under reasonable assumptions, the algorithm generates a sequence of iterates whose first-order stationary measure diminishes in expectation. In addition, we identify the iteration and sample complexity for obtaining a first-order ϵ -stationary iterate in expectation. The results of numerical experiments demonstrate the efficiency and efficacy of our proposed algorithm compared to a penalty method and an augmented Lagrangian method.

4 - Algorithmic Approaches for Identifying the Trade-off Between Pessimism and Optimism in A Stochastic Fixed Charge Facility Location Problem

Man Yiu Tsang, Texas Tech University, Lubbock, TX, United States, Karmel Shehadeh

We present algorithmic strategies for identifying the trade-off (TRO) between adopting a distributional belief and hedging against distributional ambiguity for modeling demand uncertainty in a capacitated fixed charge facility location problem (CFLP). First, we introduce a TRO model for the CFLP (TRO-CFLP), which finds the optimal number and location of facilities that minimize the fixed cost of establishing them and the maximum expected operational costs evaluated over all distributions residing within a TRO set. The TRO set is characterized by the empirical distribution, an ambiguity set, and a TRO parameter that controls the trade-off between solving the problem under the empirical distribution and the worst-case distribution within the ambiguity set. Our TRO-CFLP model enables decision-makers to explore a spectrum of location decisions, from optimistic to conservative ones. Second, we propose a Spectrum Search Algorithm to identify the full spectrum of optimal solutions for the TRO-CFLP. Third, we propose a decomposition framework for solving the TRO-CFLP model with a fixed TRO parameter. This framework is employed as a subroutine in the Spectrum Search Algorithm and includes a new hybrid primal-dual column-and-constraint generation (hC&CG) algorithm. Our numerical results demonstrate the efficiency of the Spectrum Search Algorithm and the superior computational performance of hC&CG over state-of-the-art decomposition algorithms. Our results also highlight the practical benefits of adopting solutions on the TRO-CFLP's spectrum over those obtained using traditional models. While the proposed algorithms are motivated by the TRO-CFLP problem, they can be tailored to other applications.

5 - Inventory Design for Mobile Food Markets Considering Customer Purchasing Behavior

Zijian Wang, University of Minnesota, Minneapolis, MN, United States, Yiling Zhang, William Cooper

Food insecurity, driven in part by limited access to grocery stores, affects millions across the U.S., including over 9% of Minnesotans. Mobile markets, which bring affordable and healthy foods directly to underserved communities via vehicles, have emerged as a promising solution to alleviate this issue. However, many mobile markets struggle with operational inefficiencies and financial sustainability.

In this study, we investigate how to optimize mobile market operations by focusing on inventory decisions and their relationship to customer purchasing behaviors. We model this as a bilevel optimization problem, where the mobile market operator chooses the inventory to load onto the vehicle, and sequentially visited customers make purchasing decisions based on product availability and budget constraints.

To address the computational challenges of this bilevel model, we develop an exact reformulation and propose a heuristic approach that balances solution quality and efficiency. Our numerical experiments demonstrate the effectiveness of these methods and provide insights into inventory planning for mobile markets. Numerical experiments also demonstrate the robustness of our methods in diverse scenarios.

This work offers a novel decision-making framework for mobile market operators, helping to improve service equity and operational efficiency. Future work will extend this framework to incorporate additional factors, such as routing decisions, pricing strategies, and stochastic customer behavior.

TE38

Building B Level 2 B212

Optimization, Machine Learning, and Beyond

Invited Session

OPT: Computational Optimization and Software

Chair: Shuvomoy Das Gupta, Rice University, Houston, TX, United States

1 - Tradeoffs Between Memory and Oracle-Complexity in Convex Optimization

Moïse Blanchard, Columbia University, New York, NY, United States

How much memory do convex optimization algorithms need? Algorithms for first-order convex optimization are usually divided into two categories: gradient-descent-based algorithms and cutting plane algorithms. While gradient descent methods only need to store a few vectors, which requires $\tilde{O}(d)$ bits, they require $1/\epsilon^2$ gradient queries to reach an ϵ -approximate solution. On the other hand, cutting plane methods converge exponentially: they only require $O(d \log 1/\epsilon)$ queries in dimension d . However, these are significantly more memory intensive since they require storing a matrix with d^2 bits. Towards understanding this memory/query-complexity landscape, a natural question is whether the optimal convergence rate for low-memory algorithms should always be polynomial (query-complexity polynomial in $1/\epsilon$) or can be exponential (query-complexity logarithmic in $1/\epsilon$). We provide memory/query-complexity lower bounds for the related feasibility problem, in which the goal is to find a point in a convex feasible set with access to a separation oracle. In particular, we show that deterministic algorithms always require a polynomial query-complexity in $1/\epsilon$ if they have less than quadratic memory in d . This reveals a sharp phase transition since cutting plane methods only require $O(d \log 1/\epsilon)$ queries using quadratic memory in d . To the best of our knowledge, these are the first exponential memory/query-complexity lower bounds in continuous optimization.

2 - Bridging Convex and Nonconvex Approaches: Geodesic Strong Convexity of Qcqp's from Tightness and Strict Complementarity of Sdr

Jinxin Wang, The University of Chicago, Chicago, IL, United States, Huikang Liu, Chonghe Jiang, Anthony So

Nonconvex quadratically constrained quadratic programs (QCQPs) have received increasing attention in recent decades. Two typical approaches to tackling such challenging problems are convex semidefinite relaxation (SDR) and nonconvex iterative methods (e.g., gradient descent). Although the convex SDR and nonconvex approaches have been extensively studied, they have been largely explored in parallel, and few things can be said about their connections. This paper takes the first step in building connections between convex and nonconvex approaches. Under the tightness and strict complementarity of SDR, we find that nonconvex QCQPs admit nice geometric properties, including quadratic growth, error bound, and (quotient) geodesic strong convexity. These structured properties quantify the growth behavior of optimization problems around global optimal solution sets and can imply fast convergence rates of various nonconvex iterative methods.

3 - Optimal Convergence Rates of Alternating Gradient Descent-Ascent in Bilinear Games

Shuvomoy Das Gupta, Rice University, Houston, TX, United States

We study the alternating gradient descent-ascent (AltGDA) algorithm in two-player zero-sum games. Alternating methods, where players take turns to update their strategies, have been recognized as simple and practical approaches for learning in games, exhibiting much better numerical performance than their simultaneous counterparts. However, the theoretical understanding of alternating algorithms remains limited---either restricted to the unconstrained setting or with suboptimal convergence rates. Notably, the behavior of AltGDA in constrained settings (e.g., over the probability simplex) has not been studied. We prove that AltGDA leads to constant regret for each player, and thus an optimal $O(1/T)$ ergodic convergence rate in terms of the duality gap. Our proof has several interesting departure points from classical proofs of convergence in saddle-point settings. Notably, we are unable to conclude our result from a simple "descent inequality", and must instead reason across iteration intervals that are defined by when the dynamics touch the relative boundary of the simplices. While our eventual proof is purely theoretical, we arrived at our hypothesis that AltGDA achieves $O(1/T)$ convergence via a computer-assisted methodology called the performance estimation problem (PEP) framework.

4 - Enhancing Average-Reward Reinforcement Learning via Anchoring

Jongmin Lee, Seoul National University, Seoul, Korea, Republic of, Ernest Ryu, Cominetti Roberto, Mario Bravo

Anchoring-based acceleration, first introduced by Halpern, has been actively studied in the fixed-point iteration and minimax optimization literature. More recently, anchoring acceleration has shown promise in Reinforcement Learning (RL) setups, particularly when applied to Value Iteration, one of the foundational RL algorithms. In this talk, we will introduce Anchored Value Iteration and discuss how it improves convergence and sample complexity across various average-reward RL settings, including the tabular setup, the generative model setup, and offline RL with function approximation.

TE39

Building B Level 2 B213

Supply Chain Applications of Large-Scale Discrete Optimization

Invited Session

OPT: Integer and Discrete Optimization

Chair: Madison Van Dyk, Amazon, Bellevue, WA, United States

Co-Chair: Cristiana Lara, Amazon, Bellevue, WA, United States

1 - Algorithms for Hub Location with Network Connectivity Constraints

Ozgun Elci, Amazon, Seattle, WA, United States

In transportation and logistics networks, hub location problems arise when it is necessary to route flows between origin-destination pairs through a set of hub facilities. We study developing algorithms to solve large-scale hub location problems. This problem involves selecting the best locations for hubs to connect origin-destination pairs. We consider a variant with network connectivity constraints in which we minimize the total weighted distance subject to capacity constraints on each hub, ensuring network connectivity between hubs, aiming to balance transportation costs and hub congestion. The mixed-integer programming formulation becomes computationally intractable as the number of candidate hub locations increases. We explore both exact optimization techniques and heuristic approaches to efficiently solve the underlying hub location problem.

2 - The Single-Vehicle Routing Problem with Precedence Constraints

Qie He, Amazon, Santa Clara, CA, United States, Michael Rice, Renato Werneck

The single-vehicle routing problem is a fundamental model in transportation and logistics. Many practical applications impose additional constraints, such as precedence constraints among nodes. We present an efficient heuristic to solve this problem.

3 - Large Scale Resource Allocation: a Column Generation Multi-Device Gpu Based Algorithm

Marcus Poggi, Amazon, Bellevue, WA, United States, Qi Chen, Ozlem Bilginer, Granville Paules, Zsolt Csizmadia

Resource allocation problems involve a set of tasks, each with a list of fulfillment options that consume available resources at varying levels. The objective is to minimize the sum of option assignment costs plus a linear function of resource usage. Our application context requires both primal allocations and dual lower bounds within strict time constraints—one or two orders of magnitude faster than commercial solvers need to compute linear relaxations. To meet these requirements, we employ preprocessing techniques alongside dual and primal algorithms to hot-start a Dantzig-Wolfe decomposition of the original problem. Our Column Generation algorithm coordinates agents distributed across multiple GPU devices that simultaneously solve the master linear programming problem, address column generation subproblems, search for improved primal allocations, and enhance dual bounds. We present computational results for instances involving tens of millions of tasks, demonstrating the evolution of primal and dual bounds over time. Our approach outperforms existing alternative methods.

4 - Gpu Acceleration of First Order Methods for Large Scale Resource Assignment - Benchmark Platform and Results

Qi Chen, Amazon, Bellevue, WA, United States, Ozlem Bilginer, Marcus Poggi, Zsolt Csizmadia, Granville Paules

While GPU acceleration has transformed predictive analytics, its widespread application to prescriptive analytics (optimization) is only recently emerging. At Amazon, large scale capacitated resource assignment optimization problems are prevalent across multiple business domains, yet often exceed commercial solver scalability limits. This has motivated competing implementations of various first order methods. We present a vectorized modeling approach for resource assignment problems designed for GPU accelerated first order algorithms. Using our new internal benchmark platform, we demonstrate significant performance improvements, including a 700x runtime reduction against the equivalent vectorized CPU implementation for one production instance. Compiling a library of internally-relevant instances, we evaluate algorithm sensitivity to various tuning parameters. Additionally, we demonstrate extensibility of our modeling approach to nested assignment structure. Our results build trust in the practicality of GPU accelerated optimization for business-relevant scale instances.

5 - Scenario-Optimal Recourse Policy for the Vehicle Routing Problem with Stochastic Demands

Matheus Ota, University of Waterloo, Waterloo, ON, Canada, Ricardo Fukasawa

The Vehicle Routing Problem with Stochastic Demands (VRPSD) is a well-studied variant of the Capacitated Vehicle Routing Problem in which customer demands are random variables, and the objective function accounts for the expected recourse cost. Existing exact algorithms for the VRPSD rely on strong assumptions about the probability distribution of the demands to ensure computational tractability. In contrast, the general stochastic programming community often models uncertainty using a finite set of scenarios, enabling the approximation of more realistic and diverse probability distributions through sampling. In this work, we propose a *scenario-optimal recourse policy* that assumes perfect information of the demand realization under each scenario. Leveraging known results from network-flow and polyhedral theory, we show that the resulting recourse function can be fully described by certain cuts that resemble the well-known rounded capacity inequalities (RCI's). This scenario-optimal recourse policy is then used to under-approximate the classical recourse policy. Computational experiments show that our method yields stronger bounds than an approach that is based on an extension of a recently proposed integer L-shaped-based method.

TE40

Building B Level 2 B214

Network Performance, Disruption, and Adaptation

Contributed Session

Contributed

Chair: Chanchal Kumar Salode, Indian Institute of Technology Delhi, Department of Management Studies, New Delhi, 110016, India

1 - Equitable and Stochastic Humanitarian Relief Network Design for Disaster Preparedness

Chun Peng, Beijing Jiaotong University, Beijing, China, People's Republic of

In this study, we propose an equitable stochastic humanitarian relief network design problem for disaster preparedness when the demand is uncertain. Specifically, we develop an inequity-averse aggregation function to capture the equity by balancing humanitarian benefits and allocation efficiency. Using sample average approximation technique, we reformulate the two-stage stochastic model as a large-scaled optimization problem, which can be further efficiently solved by Benders decomposition algorithm with accelerating strategies. We finally demonstrate the value of our methodology in mitigating inequities using both synthetic and real-life data.

2 - Flight Delay Dynamics: Unraveling the Impact of Airport-Network-Spilled Propagation on Airline on-time Performance

Yi Tan, The University of Alabama in Huntsville, Huntsville, AL, United States, Yajun Lu, Lu Wang, Hieu Pham

Flight delay prediction has attracted increasing attention in airline operations. Early identification of potential flight delays is crucial for improving airport scheduling and airline operations while mitigating associated costs. This study investigates the influence of the potential propagation of flight delays throughout the airport network via interconnected flights, a mechanism we term Airport-Network-Spilled Propagation (ANSP). To model the ANSP mechanism, we develop a novel time-dependent, network-based approach that decays the importance of past delays. From this network, we extract a real-time ANSP score for each airport to measure the influence of propagated delays. To evaluate our proposed approach, we employ four state-of-the-art machine learning models using domestic airline on-time performance data from the 30 Large Hub airports in the United States. The results demonstrate that integrating the ANSP score with established features from airline operations literature significantly enhances flight departure delay prediction performance, achieving an increase in AUC of up to 5.49%. Furthermore, we conduct an explainable AI analysis using Shapley additive explanations (SHAP), which reveals that our ANSP score ranks as the most important predictor among all features tested.

3 - Sparsity-Induced Gradient-Free Optimization of Heterogeneous Information Perturbations in Transportation Networks

Shiwei Liu, University of Florida, Gainesville, FL, United States, Hongcheng Liu, Lili Du

Urban navigation platforms can alleviate traffic congestion by subtly altering the travel-time information they broadcast, thereby nudging drivers toward routes that benefit the overall network performance. This paper proposes a Heterogeneous Perturbation Optimization Model (HPOM), which assigns a distinct perturbation weight to each link within a Coordinated Routing Mechanism with Mixed-strategy Information Perturbation (CRM-M-IP). The objective of HPOM is to minimize total system travel

time while explicitly constraining both the average and worst-case user sacrifice. We introduce a two-stage solution framework to solve HPOM. Stage 1 employs a self-adaptive mirror descent (MD) routine to compute each mixed-strategy perturbed equilibrium routing decision (MP-ERD) in real time, with a convergence rate of $O(1/K)$. Stage 2 incorporates a Sparsity-Inducing Stochastic Gradient-Free (SI-SGF) algorithm to solve HPOM with a dimension-insensitive $O(1/K)$ convergence rate for convex objectives, and it extends gracefully to non-convex settings. A golden-section search algorithm further accelerates convergence, making the approach suitable for real-time deployment. Comprehensive experiments on the Sioux Falls and Chicago networks demonstrate that the proposed MD algorithm outperforms simultaneously-updating distributed algorithm (SDA) in solving MP-ERD, particularly when high convergence precision is required or when the problem scale is large. Furthermore, compared to the no-perturbation baseline, the heterogeneous perturbation framework reduces total travel time by up to 25.67% more than the homogeneous approach, while keeping user sacrifice low.

4 - Single Allocation Hub Line Location Problems: Model Formulations and Exact Solution Methods

Chanchal Kumar Salode, Indian Institute of Technology Delhi, New Delhi, India, Prasanna Ramamoorthy

In this study, we introduce the Single Allocation Hub Line Location Problem, which focuses on locating a set of hubs connected in the form of a line, while assigning non-hub nodes to hubs under a single allocation scheme. This problem finds applications in various domains such as public transportation, telecommunications, and rapid transit systems, where the cost of designing networks is high compared to routing costs, making full interconnection of hubs impractical. The objective is to minimize the total transportation cost between all pairs of origin-destination nodes. To address this problem, we propose path-based and flow-based Mixed Integer Linear Programming (MILP) models. Furthermore, to efficiently solve large-scale instances, we develop a Benders Branch-and-Cut algorithm incorporating state-of-the-art enhancements. Computational results on two sets of benchmark instances demonstrate the efficiency of the proposed algorithm, which significantly outperforms a general-purpose solver in both speed and scalability.

TE41

Building B Level 3 B301

Data and Equity in Clinical Decisions

Invited Session

Health Applications Society

Chair: Anqi Wu, Florida International University, Miami, FL, United States

Co-Chair: Chunjie Zhao, Clark University, 950 MAIN STREET, Worcester, MA, 01610-1477, United States

1 - Broken Records, Broken Products: How Data Integrity Shapes Recall Risk

Chunjie Zhao, Clark University, Worcester, MA, United States, Anqi Wu

Product recalls represent critical events that can severely disrupt business operations, damage brand reputation, and impose substantial financial costs. Despite extensive research on the operational and regulatory aspects of product recalls, the role of data integrity violations, defined as breaches in data accuracy, consistency, and reliability, remains underexplored in the literature on information systems. This study investigates the impact of data integrity violations on the timing and likelihood of product recalls. Using survival analysis in a longitudinal data set of product recall events, we find that products linked to data integrity violations face a 13% higher risk of being recalled compared to those affected by other factors. This risk increases to approximately 15% within product quality-focused project domains. These findings highlight the critical role of data integrity in ensuring product safety and organizational resilience. The study contributes to the literature by establishing an empirical connection between data governance and operational risk and offers actionable insights for firms aiming to enhance product quality and recall prevention through improved data management practices.

2 - Reducing Transparency? the Hidden Gender Cost of Ehr Adoption in Organ Discards

YUYING CAI, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, Kai-Lung Hui

The U.S. organ transplant system faces a dual crisis: a persistent organ shortage coexists with high discard rates, despite many unused organs being of comparable quality to those successfully transplanted. Gender disparities further compound this inefficiency, as organs from female donors, who represent a larger share of deceased donors, are disproportionately discarded. This study examines whether Electronic Health Record (EHR) adoption in donor hospitals mitigates or exacerbates gender disparities in organ discards, particularly through its spillover effects on Organ Procurement Organizations (OPOs). Using hospital-level EHR adoption data from the AHA IT Supplement and individual-level deceased donor records from UNOS (1987–2019), we employ a generalized difference-in-differences (DiD) design to address selection bias. Contrary to the expectation that EHRs improve transparency and equity, we find that EHR adoption widens the gender gap in discard rates, with organs from female donors facing significantly higher discard probabilities. We further conduct a series of robustness checks to confirm this finding. Our study contributes by revealing unintended consequences of EHRs, challenging assumptions about their uniformly positive effects. It also bears immediate relevance for Health Information Exchange (HIE) policymakers and transplant stakeholders aiming to ensure that digital technologies promote, rather than hinder, equitable healthcare delivery.

3 - Bridging the Digital Divide: An Experimental Investigation of Telehealth Adoption in Pakistan

Edwin (Yuchen) Wang, University of Massachusetts Boston, Boston, MA, United States, Adeela Gulzari

While digital health platforms hold promises for expanding access to care, their effectiveness hinges on users' ability to adopt and trust these systems, especially in low-resource and rural contexts. We designed an experiment to investigate the key barriers to online healthcare adoption among countryside residents in Pakistan. We recruit adult participants and randomly assign them to one of two versions of a prototype telemedicine application: (1) a basic interface offering text-only consultations and appointment scheduling, and (2) an enhanced version featuring video consultations, interactive health education content, and a peer-support forum. We hypothesize that richer media features and social support tools will significantly increase trust and perceived usefulness, thereby improving adoption rates. Our findings will contribute to the imbalance of healthcare resources by (i) isolating context-specific design principles for telehealth in developing regions, (ii) empirically validating the roles of trust and media richness in digital-health uptake, and (iii) offering actionable guidelines for policymakers and practitioners seeking to bridge rural healthcare gaps through information technology.

4 - From Deception to Perception: the Surprising Benefits of Deepfakes for Detecting and Measuring Nuanced Bias Patterns in Facial Regions

Yizhi Liu, University of Maryland, College Park, College Park, MD, United States, Balaji Padmanabhan, Siva Viswanathan

While deepfake technologies are often criticized for their misuse, we demonstrate their surprising utility for bias detection and measurement in high-stakes decision-making contexts. We develop a novel methodological framework that uses deepfake-generated facial images to precisely manipulate sensitive attributes (e.g., race, age) while holding all other visual features constant. Applying this approach to pain assessment, a domain where subjective interpretation of facial expressions influences clinical decisions, we uncover systematic patterns of bias across assessor groups. Using a series of crowdsourced experiments, we find that racial and age-based biases are not only prevalent but also region-specific. In particular, White assessors show greater bias in upper facial regions, while Black assessors exhibit stronger bias in lower facial regions. This AU-level analysis offers unprecedented granularity in understanding the perceptual mechanisms behind biased decision-making. Our findings reveal how deepfake-driven visual correspondence studies overcome longstanding limitations in experimental control, enabling more accurate and fine-grained bias measurement. This work not only advances the methodological frontier of bias research but also offers practical implications for designing fairer human and AI-based assessment systems in visual decision-making contexts.

TE42

Building B Level 3 B302

Drug and Treatment Development: The Role of Clinical Trials, Post-Approval Data, and Incentives

Invited Session

Health Applications Society

Chair: Ozge Yapar, Indiana University, Bloomington, IN, United States

1 - Optimal Push, Pull, and Failure Funding for Global Health

Chenxi Xu, Duke University, Durham, NC, United States, David Ridley, Peng Sun

Malaria caused over 600,000 deaths in 2021, yet commercial incentives are weak for drug and vaccine development for malaria and other tropical diseases. Governments and nonprofits address these market failures through push (e.g., grants) and pull (e.g., prizes). We propose a third mechanism in which the funder only pays when the firm fails. That is, the funder reimburses the firm a share of its testing costs, which we refer to as "insurance." Insurance is optimal when the commercial market is large enough to reward success, but too small to induce investment. The optimal mechanism addresses adverse selection and moral hazard. For most tropical diseases, including malaria, we recommend pull funding with supplementary push support. Insurance is optimal for tuberculosis if testing costs are less than a billion dollars. These findings challenge current practices dominated by push funding and extend to funding innovations in other sectors.

2 - Clinical Trials and Conditional Approval Schemes: Implications for Healthcare Systems and Companies

Ozge Yapar, Indiana University, Bloomington, IN, United States, Andres Alban

Traditionally, companies complete Phase III trials before applying for a healthcare payer's approval. Conditional approval schemes allow earlier market entry based on preliminary evidence, with full approval contingent on further data demonstrating health-economic value. The use of these schemes alters the incentives around the design of trials by decoupling market entry from trial completion.

These schemes may encourage companies to enroll more patients, generating stronger evidence for the new treatment. However, companies choose how much evidence to generate before applying for conditional approval. Some may apply early with limited data, speeding up patient access but increasing uncertainty. Others may wait for more interim data, delaying patient access and making comparisons to traditional timelines less clear.

We use a Bayesian decision-theoretic model in which a company chooses the trial's sample size and timing of an interim analysis based on anticipated outcomes of cooperative bargaining with a payer. Negotiations at interim and final analysis determine approval decision and pricing. These negotiations are modeled with a Nash cooperative bargaining model in which the company maximizes profits, and the payer maximizes population health value.

We find conditional approval schemes increase the overall sample size of trials, but market entry timing depends on the company's beliefs about expected profit.

3 - Underrepresentation Bias in Clinical Trials: the Ripple Effect on Medical Decision-Making and Patient Outcomes

Xiaoshan Peng, Indiana University, Bloomington, IN, United States, Ling Li, Kurt Brethauer

Failures of health equity are closely intertwined with biases in healthcare treatment, such as bias due to race, ethnicity, gender, and age. This paper studies how underrepresentation bias in clinical trials affects physician decision-making, patient outcomes, and equity. We propose a behavioral model to analyze physicians' treatment decisions, which are influenced by two distinct cognitive systems: the rational and the intuitive systems. We aim to explore whether these decision-making systems exacerbate healthcare disparities by relying on biased clinical trial results or if they can adapt to mitigate underrepresentation bias. The rational system is modeled as a multi-armed bandit problem with Bayesian updating. The intuitive system is modeled as a Markov decision process using a "win-stay, lose-shift" heuristic. Physicians may alternate between the two cognitive systems, following a dual process model. We find that the impact of the underrepresentation bias depends on the effectiveness of the new treatment. If a less effective or equally effective treatment is introduced, the rational system will correct the impact of the bias and choose the optimal treatment eventually. Otherwise, the physicians' treatment decision may not converge to the optimal one. The impact is amplified by the size of the clinical trials. In contrast, the dual process model corrects the impact of the bias such that the physician eventually chooses the more effective treatment under mild conditions. Furthermore, we present a case study on how gender underrepresentation in clinical trials affects the use of low-dose aspirin for preventing myocardial infarction and ischemic stroke.

4 - Incentivizing Drug Rediscovery: Exclusivity Extension and Priority Review for Repurposed Therapeutics

Elodie Adida, University of California, Riverside, CA, United States, Fernanda Bravo

Drug repurposing consists in identifying new indications for existing drugs. Compared to de novo drug development, drug repurposing incurs less risk and costs and requires less time. However, it is currently under-utilized, primarily due to a lack of financial incentives for drug manufacturers. A new drug enjoys only a limited time with protected market exclusivity. After then, the drug is often subject to competition with a generic drug and the potential for profits drops sharply, making investment in developing new indications often cost-ineffective. While a new indication is granted a temporary exclusivity period, it is often too short to recoup investment costs. We study two mechanisms to better incentivize drug repurposing: (i) granting an extension of the exclusivity period for the primary indication upon approval of a new (secondary) indication, and (ii) offering priority in the review process to speed up approval of the secondary indication. For each mechanism, we investigate how to best design the intervention and what the effect is on the welfare of patients of each indication and the overall effect on patient welfare. We also compare the two types of intervention in terms of patient welfare. We find that while both interventions can be effective in improving the incentive to engage in drug repurposing, they may hurt the welfare of patients of the primary indication and the overall patient welfare. We find that priority review is more likely to be beneficial to patients than exclusivity extension and thus appears to be the dominant type of intervention.

TE43

Building B Level 3 B303

Emerging Technologies in Integrated Risk Management

Invited Session

MSOM: iForm

Chair: Danko Turcic, UCR, Riverside, CA, United States

Co-Chair: Ye (Angelina) Liu, Syracuse University, Syracuse, NY, United States

1 - Fragmentation and Visibility in Supply Networks

Nikolay Osadchiy, Emory University, Atlanta, GA, United States, Maxi Udenio, Vishal Gaur

Supply network visibility is a key for resilient supply chains, yet, in practice, visibility beyond tier-1 suppliers is still rare. Because of that, many firms are caught off-guard by disruptions to their upper-tier suppliers. Using the state-of-the-art firm-level buyer supplier relationship data, we construct a time evolving supply network and explore its structure, with the goal of improving visibility into the network. We find that the network is highly dynamic, rewiring about 25% of its links each year. The network is also fragmented, i.e., it contains communities of firms that are densely connected to each other yet sparsely connected to other communities via a distinct set of firms. Moreover, the fragmentation increases over time, meaning that the network has more communities and communities equalize in size. Based on this structure, we propose an efficient network exploration algorithm that enables firms to find a disrupted supplier in their networks faster. The performance of the algorithm improves as the network becomes more fragmented.

2 - Cvar Based Risk Sensitive Policy Learning

Jingyuan Wang, New York University, New York, NY, United States, Zhimei Ren, Zhengyuan Zhou

The conventional offline multi-action policy learning literature aims to find a policy that performs well in terms of the average policy effect (APE) on the population, i.e. the social welfare. However, in many settings, including healthcare and public policies, the decision-maker also concerns about the risk of implementing certain policy. The optimal policy that maximizes social welfare could have a risk of negative effect on some percentage of the worst-affected population, hence not the ideal policy. In this paper, we investigate risk sensitive offline policy learning with conditional value at risk (CVaR) of covariate-conditional average policy effect (CAPE) as a risk measure. To this end, we first provide a doubly-robust estimator for the CVaR of CAPE, and show that this estimator enjoys asymptotic normality even if the nuisance parameters suffer a slower than $1/n^{1/2}$ estimation rate (n being the sample size). We then propose a risk sensitive learning algorithm that finds the policy maximizing the weighted sum of APE and CVaR of CAPE, within a given policy class. We also show that the sub-optimality gap of the proposed algorithm is in the same order of the canonical policy learning schemes that maximizes APE. The proposed methods are evaluated empirically, demonstrating that by sacrificing not much of the social welfare, our methodology improves the outcome of the worst-affected population.

3 - Energy-Weighted Flow Matching for Offline Reinforcement Learning

Weitong Zhang, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Shiyuan Zhang, Quanquan Gu

This paper investigates energy guidance in generative modeling, where the target distribution is defined as $q(\mathbf{x}) \propto p(\mathbf{x})\exp(-\beta E(\mathbf{x}))$, with $p(\mathbf{x})$ being the data distribution and $E(\mathbf{x})$ as the energy function. To comply with energy guidance, existing methods often require auxiliary procedures to learn intermediate guidance during the diffusion process. To overcome this limitation, we explore energy-guided flow matching, a generalized form of the diffusion process. We introduce energy-weighted flow matching (EFM), a method that directly learns the energy-guided flow without the need for auxiliary models. Theoretical analysis shows that energy-weighted flow matching accurately captures the guided flow. Additionally, we extend this methodology to energy-weighted diffusion models and apply it to offline reinforcement learning (RL) by proposing the Q-weighted Iterative Policy Optimization (QIPO). Empirically, we demonstrate that the proposed QIPO algorithm improves performance in offline RL tasks. Notably, our algorithm is the first energy-guided diffusion model that operates independently of auxiliary models and the first exact energy-guided flow matching model in the literature.

4 - Optimization of Multi-Product (R,S,Q) Inventory Policies Under Aggregate Service Level Constraints for Spare Parts Management

Patrick Helm, Technical University of Munich, Munich, Germany, Stefan Minner, Gudrun Kiesmüller

Managing spare part inventories is challenging since demand is highly intermittent and replenishment decisions must be made based on very limited data. Nevertheless, high service levels must be ensured to avoid costly stock-outs.

In contrast to prior work, we consider the peculiarities of real-world spare parts demand data. Thus, we develop a classification scheme to select the right demand model for each SKU such that it generalizes well without overfitting the few available observations. Moreover, we use chance constraints to make sure that service targets are reached in a short horizon, despite the randomness of service measures. Finally, we consider demand correlation across products and time, which is habitually neglected in existing work.

We focus on multi-product (R,s,Q) inventory policies, subject to an aggregate fill rate constraint. We set up a nonconvex mixed-integer optimization problem to jointly determine the cost-optimal policy parameters for all SKUs in a segment. We develop a scalable algorithm based on Lagrangian Relaxation coupled with continuous optimization and gradient-based rounding to solve even large instances efficiently.

We validate the proposed algorithm on a real-world demand data set of an after-sales platform in the machinery industry. The algorithm scales easily to segments of more than 5000 SKUs. We demonstrate the effectiveness of the chance constraints in ensuring that fill rate targets are achieved in a short horizon. Finally, we also show the effects of neglecting demand correlation on the fill rate performance and highlight the importance of choosing the right model for demand intermittency for each SKU.

5 - Decision Simulation with LLMs: Value of Deliberation and Natural Language

Mahtab Raeisdanaei, University of Illinois Chicago, Chicago, IL, United States, Selva Nadarajah, Daniel Jiang

Large language models offer new opportunities to simulate human decision making, but most systems rely on static user profiles, symbolic features, or demographic traits. In contrast, real decisions emerge from how individuals reason about the specific structure of the problem they face. Research in decision science and cognitive psychology shows that these reasoning processes are not random but the outcome of internal deliberation that can be captured in natural language. Building on these foundations, we present a framework that uses language models to predict decisions within a specific business domain, using energy procurement as a testbed. Our approach starts from natural language reasoning traces and organizes them into structured natural language components such as risk, values, and tradeoffs based on known decision theoretic frameworks. This structured trace is used to predict the final decision. This setup captures how individuals prioritize different aspects of the same problem. Some may focus on risk, others on value alignment. By structuring reasoning rather than relying solely on demographic and problem features, we show how large language models can support decision simulation grounded in theoretical frameworks of how people think, thereby uncovering the value of capturing internal deliberation in natural language form.

TE44

Building B Level 3 B304

Inventory Management

Contributed Session

Contributed

Chair: Kun Soo Park, Seoul National University, Seoul, Korea, Republic of

1 - Joint Assortment and Inventory Optimization for the Front-End Warehouse

Kexin Wu, National University of Singapore, Singapore, Singapore, Xiaobo Li, Xue-Ming Yuan

This study examines assortment and inventory optimization in a two-warehouse e-commerce system that seeks to balance delivery speed with cost efficiency. The system consists of a front-end warehouse, strategically positioned near customers to support rapid order fulfillment but limited by capacity constraints, and a back-end warehouse, which holds a larger inventory but entails higher fulfillment costs and longer shipping times. We model this system as a joint assortment and inventory optimization problem, where initial assortment and inventory decisions are made for the front-end warehouse prior to customer demand realization. Once orders are placed, the e-commerce company must decide the quantity of each order type to fulfill from the front-end warehouse, with the remaining quantities fulfilled from the back-end warehouse. Our approach uses a two-stage stochastic programming model to minimize total fixed assortment costs and non-fulfillment costs. We derive the closed-form optimal policies for specific order structures, including distinct single-item sales with non-overlapping bundles, as well as configurations like series and star cover bundles. For more general order structures, we develop an approximation algorithm and establish a tight performance guarantee bound. Furthermore, we demonstrate that the joint assortment and inventory optimization problem is NP-hard, even for simplified scenarios with deterministic demand and single-item order types. To address this complexity, we introduce a heuristic algorithm that enables efficient assortment decision-making. We conduct the numerical experiments using simulated data to evaluate the effectiveness of our proposed policies. The results show that our solutions achieve a good balance between optimality and computational efficiency.

2 - Model Predictive Control for Inventory Management with Advanced Order Information in Supply Chain Planning

Olorunfunmi Osinubi, The University of Texas at Arlington, Arlington, TX, United States

In supply chain planning, it is very important to forecast changes in the market in order to maintain an inventory level that is just enough to satisfy customer demand. This poster presents a Model Predictive Control (MPC) framework that optimizes inventory management by incorporating advanced order information (AOI). Our case study centers on a small-scale additive manufacturing (AM) business, where production and inventory decisions are tightly constrained by limited resources. The proposed

model captures daily inventory dynamics influenced by production rates, customer product demands, and waste collection schedules. Deliveries are categorized by urgency—next-day, speed, and standard—to reflect real-world logistics practices where businesses prioritize shipments based on customer urgency and delivery deadlines. By leveraging AOI, the model anticipates future demands and adjusts production accordingly. The MPC formulation minimizes a cost function that penalizes deviations from target inventory levels and inefficient production effort. Solved iteratively over a finite prediction horizon, this approach remains responsive to system changes in real time. Overall, the proposed solution offers a scalable, adaptive strategy that enhances inventory efficiency, supports reliable order fulfillment, and strengthens planning for small AM business supply chains under demand uncertainty.

3 - A Newsvendor Approximation for Inventory Management with Shoplifting

Dincer Konur, Texas State University, San Marcos, TX, United States

Shoplifting has been on the rise and it makes inventory management even more challenging. In this study, we investigate a single-product single-season inventory control in the presence of shoplifters. First, we conduct detailed simulation and analyze the impact of shoplifting on order decisions and profits. Then, a novel newsvendor model is formulated to approximate the simulation outcomes. The model and simulation give similar results and produce parallel insights about the effects of shoplifting for seasonal inventory management. Particularly, we observe that shoplifting tends to increase the optimal order quantity while decreasing the profits. We also discuss how to use the model to quantify the economic impact of shoplifting beyond the direct losses due to shoplifted amount. Finally, we extend the model to consider optimum levels of investments in different shoplifting countermeasures.

4 - Analyzing Inventory Models for Maintenance and Repair Operations (Mro)

Sundaravalli Narayanaswami, Indian Institute of Management, Ahmedabad, India

Inventory management holds great importance in effective running of core industries. MRO inventory is particularly challenging because of its importance in routine plant operations and the uncertain nature of the demand. Organizations often face a trade-off: high inventory levels tie down working capital and the products can become Non-Moving inventory (NMI) due to technology obsolescence while low inventory could lead to stock-outs affecting production.

This study aims at improving the MRO inventory control through the application of more sophisticated inventory control techniques like the Just-In-Time (JIT) and the Economic Order Quantity (EOQ). The objective is to reduce holding and procurement cost and guarantee that the required MRO inventory is available.

An extensive literature on current inventory management practices is conducted. Inventory models were evaluated by using real life data to understand their relevance in managing MRO inventory. Typically, organizations adopt an ABC (based on volume) or FSN (Fast-, slow-, non- moving) categorization of inventory as a standard inventory management practice. Using Poisson distribution to model uncertainty of demand, we develop a Pareto optimization model to minimize inventory costs, improve availability of inventory, and to better decision-making through categorization and analysis. As a testbed, we have considered organizations from oil industry and automobile manufacturing. Our findings indicate that a VED (Vital, Essential, Desirable) approach is more suited than current practices in reducing financial inefficiencies and to enhance operational sustainability by addressing risks associated with NMI. Results of our analysis are reported.

5 - Bandit Algorithm for Online Learning in Inventory Problems with Lead TIMES

JeongWook Lim, Seoul National University, Seoul, Korea, Republic of, Byeongkwon Lee, Kun Soo Park

We study data-driven inventory management problems using a bandit algorithm. While many data-driven approaches have been successfully applied to inventory problems, there is a lack of truly online and universal algorithms. While prior work explored deep reinforcement learning approaches as a universal methodology for inventory problems, they lack the online learning capability that bandit algorithms offer. To address issues with applying bandit algorithms to inventory problems, our algorithm has following three key concepts: 1) low-risk exploitation through a virtual inventory system, 2) management of a candidate arm set, 3) efficient exploration via Thompson sampling. We evaluate the performance of our algorithm and obtain regret upper bounds.

TE45

Building B Level 3 B305

Capacity Management and Service Design for Emerging Platforms

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Sheng Liu, University of Toronto, Toronto, ON, Canada

1 - Should I Continue to Search? Mechanism Design for TIME-Sensitive Freight Matching Platform Auctions

Ziyan Wei, Tilburg University, Tilburg, Netherlands, Chong Zhang, Jan Fransoo, Pieter Kleer

The rapid growth of digital freight matching platforms—such as Uber Freight (U.S.), Sennder (Europe), and Full Truck Alliance (China)—is evident. These platforms operate under unique time pressure: shippers face strict shipment deadlines and carriers seek to fill limited capacity quickly. However, current platform mechanisms often overlook the complexities of decision-making under such time-sensitive constraints.

This study investigates shipper's behavior across three search modes: passive (waiting for platform bids), active (seeking outside options), and hybrid (multi-homing or alternating). Motivated by real-world practices, we model the platform as a first-price sealed-bid auction with a finite deadline. Carriers may withdraw bids any time and shippers may terminate the auction upon accepting an offer. A central challenge is shipper's declining probability of successful recall—that is, the chance a previously seen bid remains available—which we model as decreasing with both time and bid value.

We characterize shipper's search process and identify a reservation utility or threshold as the optimal strategy. Our analysis confirms that the reservation utility increases with the probability of recall, consistent with prior work. However, we extend the literature by introducing an availability assurance mechanism that requires carriers to hold bids open for a minimum duration. We note that the mechanism may impact bid arrivals and values. We establish conditions under which the mechanism is beneficial for the platform to reduce multihoming and derive the optimal mechanism length.

Our findings contribute to auction theory and platform design by offering insights into search behavior, auction dynamics with recall uncertainty under time constraints.

2 - On TIME, Every Time: Integrating Autonomous Vehicles in Food Delivery

Neda Mirzaeian, The University of Texas at Dallas, Richardson, TX, United States, Soraya Fatehi, Milind Dawande, GANESH JANAKIRAMAN

Food-delivery platforms traditionally rely on crowdsourced human-driven vehicles (HVs), but leased autonomous delivery vehicles (ADV)s are emerging as an alternative. A major challenge for such platforms is keeping costs low while meeting tight delivery time commitments. We address this challenge by analyzing how to optimally manage HV and ADV fleets and whether a hybrid fleet benefits the platform, even if ADVs become more cost-effective than HVs. We develop three robust optimization models: Pure HV (using only HVs), Pure ADV (using only ADVs), and Hybrid (using both). The hybrid model enforces delivery time constraints separately for HVs and ADVs, preventing the optimal hybrid solution from being trivially superior to the pure-optimal solutions. First, in the pure settings, we determine optimal capacity levels (number of HVs/ADV)s and compensation rates to HVs (impacting their availability) and ADVs (affecting the ADV supplier's profitability from leasing its ADVs). Building on these solutions, we optimize the hybrid setting by determining HV and ADV capacity levels and compensation rates, along with demand allocation between HVs and ADVs. By comparing the three optimal solutions, we identify when the hybrid setting is optimal and estimate resulting cost savings through a calibrated numerical study. We extend these models to a non-stationary case with an off-peak and a peak period. We optimally adjust the HV supply in each period while maintaining the same ADV level (as leases span much longer periods than the peak and non-peak periods). Our calibrated numerical study then uncovers the cost advantage of the optimal hybrid strategy.

3 - Capacity Configuration in Non-Stationary Environments: the Operational Value of Prediction and Dynamic Decision-Making

Piao Hu, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Jiashuo Jiang, Guodong Lyu, Hao Sun

We study how volume flexibility, when properly configured into production systems, can be used to determine the optimal capacity level while satisfying designated performance targets, particularly when the demand is not only random but also exhibits variations in its distribution across the planning horizon. To address this challenge, we develop an online two-stage stochastic optimization framework over a finite planning horizon. During each period of the planning horizon, the planner makes the first-stage decision for capacity configuration prior to demand realization, and then allocates the capacity to fulfill multiple performance targets at the second stage when the demand is realized.

We propose an adversarial learning approach to iteratively solve the capacity configuration and allocation problem in an online manner. Our main results show that this approach provides a provably near-optimal performance guarantee while offering great computational efficiency. Furthermore, we demonstrate the value of volume flexibility in capacity configuration to minimize potential losses, particularly when the underlying stochastic inputs follow non-stationary distributions. This extends previous studies that focused on stationary inputs and paves the way for establishing a transparent data-driven oracle that can be applied to a wide range of two-stage online decision-making problems. Through extensive numerical experiments, we underscore the benefits of demand distribution information in non-stationary environments and showcase how volume flexibility, in addition to the classic process flexibility, can help mitigate the misalignment between supply and demand. Furthermore, we apply these techniques and insights to address a crowdsourcing walker deployment problem arising in last-mile delivery operations.

4 - Division of Labor and the on-Demand Economy: Implications for System Design

Milind Sohoni, University at Buffalo, The State University of New York, Buffalo, NY, United States

We study service systems in the on-demand economy that rely on a flexible workforce. Server reliability is endogenously influenced by system parameters or design choices through the division of labor. Using a queueing-theoretic framework, we show how behavioral factors shape server reliability and specialization, thereby impacting optimal capacity planning and system design.

5 - Transportation Service Region Design Under Distributional Uncertainty

Yidi Miao, Carnegie Mellon University, Pittsburgh, PA, United States, Eojin Han, Peter Zhang

We propose a novel framework for the design of stochastic clearing systems, aimed at optimizing waiting and traveling costs through the partitioning of the service region. By leveraging the accumulated operational data over time, our framework assumes that dedicated regions for vehicles can be effectively managed. To address the inherent uncertainty in demand realization, we adopt a data-driven distributionally robust optimization approach, preventing decisions from out-of-sample effects. Our primary objective is to partition the entire service area into several contiguous districts, while minimizing overall waiting and traveling costs under distributional uncertainty. This is achieved through a discrete approximation of the worst-case distribution within the ambiguity set, coupled with contiguity constraints on partitioning decisions to ensure practical implementation. The proposed framework is robust and adaptable, making it applicable to similar transportation and logistics problems.

TE46

Building B Level 3 B306

Advances in Causal Effect Estimation and Decision Making

Invited Session

Decision Analysis Society

Chair: Justin Whitehouse, Stanford University, Pittsburgh, PA, United States

Co-Chair: Vasilis Syrgkanis, Stanford University, Stanford, CA, United States

1 - A Meta-Learner for Heterogeneous Effects in Difference-in-Differences

Hui Lan, Stanford University, Stanford, CA, United States, Haoge Chang, Eleanor Dillon, Vasilis Syrgkanis

We address the problem of estimating heterogeneous treatment effects in panel data, adopting the popular Difference-in-Differences (DiD) framework under the conditional parallel trends assumption. We propose a novel doubly robust meta-learner for the Conditional Average Treatment Effect on the Treated (CATT), reducing the estimation to a convex risk minimization problem involving a set of auxiliary models. Our framework allows for the flexible estimation of the CATT, when conditioning on any subset of variables of interest using generic machine learning. Leveraging Neyman orthogonality, our proposed approach is robust to estimation errors in the auxiliary models. As a generalization to our main result, we develop a meta-learning approach for the estimation of general conditional functionals under covariate shift. We also provide an extension to the instrumented DiD setting with non-compliance. Empirical results demonstrate the superiority of our approach over existing baselines.

2 - Orthogonal Causal Calibration

Justin Whitehouse, Stanford University, Stanford, CA, United States, Steven Wu, Vasilis Syrgkanis, Bryan Wilder, Christopher Jung

Estimates of heterogeneous treatment effects such as conditional average treatment effects (CATEs) and conditional quantile treatment effects (CQTEs) play an important role in real-world decision making. Given this importance, one should ensure these estimates are calibrated. While there is a rich literature on calibrating estimators of non-causal parameters, very few methods have been derived for calibrating estimators of causal parameters, or more generally estimators of quantities involving nuisance parameters. We develop general algorithms for reducing the task of causal calibration to that of calibrating a standard (non-causal) predictive model. We study a notion of calibration defined with respect to an arbitrary, nuisance-dependent loss ℓ , under which we say an estimator θ is calibrated if its predictions cannot be changed on any level set to decrease loss. For losses ℓ satisfying a condition called universal orthogonality, we present a simple algorithm that transforms partially-observed data into generalized pseudo-outcomes and applies any off-the-shelf calibration procedure. For losses ℓ satisfying a weaker assumption called conditional orthogonality, we provide a similar sample splitting algorithm that performs empirical risk minimization over an appropriately defined class of functions. Convergence of both algorithms follows from a generic, two term upper bound of the calibration error of any model. We demonstrate the practical applicability of our results in experiments on both observational and synthetic data. Our results are exceedingly general, showing that essentially any existing calibration algorithm can be used in causal settings, with additional loss only arising from errors in nuisance estimation.

3 - Partial Identification for Long-Term Causal Inference

Xiaojie Mao, Tsinghua University, Beijing, China, People's Republic of, Songliang Chen

Long-term causal inference studies the treatment effect of a certain intervention (e.g., a product design or a therapy) on some long-term outcome of interest (e.g., users' long-term satisfaction or patients' long-term health). This problem is very challenging: randomized experiments are gold-standard for causal inference but they are often expensive and have short durations, so long-term outcome observations may not be available; observational studies can be cheaper and more likely to collect observations for long-term outcomes, but they are susceptible to confounding bias. Existing literature proposes to address this challenge by combining experimental and observational data and leveraging their complementary strengths, but this combination requires various important assumptions to achieve exact identification. We show that some common assumptions such as the well-known statistical surrogacy condition and external validity condition are prone to violation. To achieve robust long-term causal inference, we propose to relax these assumptions and derive the corresponding partial identification bounds, under both a nonparametric model and a semiparametric partially linear model. We also provide estimation and inference methods for these partial identification bounds.

4 - Causal Inference with High-Dimensional Treatments

Edward Kennedy, Carnegie Mellon University, Pittsburgh, PA, United States

In this work we consider causal inference when the number of treatment levels is comparable to or larger than the number of observations. This setting brings two unique challenges: first, the treatment effects of interest are a high-dimensional vector rather than a low-dimensional scalar and, second, positivity violations are unavoidable. We first discuss fundamental limits of estimating effects in such settings, showing consistent estimation is impossible without further assumptions. We go on to propose new

doubly robust thresholding/penalized estimators, and analyze them under exact and approximate sparsity assumptions. We then derive minimax lower bounds to characterize optimal rates of convergence. Finally we illustrate the methods in an education application studying school effects on test scores, where the number of treatments (schools) is in the thousands. This is joint work with Patrick Kramer (CMU) and Isaac Oppen (RAND).

TE47

Building B Level 3 B307

Optimization Topics II

Contributed Session

Contributed

Chair: Anna Deza, University of California-Berkeley, Berkeley, CA, United States

1 - Nationwide Cash Forecasting and Optimization System for Multi-Currency Banking Operations

Wilder Avalos, Banco de Crédito del Perú, Lima, Peru, Jan Nauta

We deployed a nationwide cash supply forecasting and optimization system for BCP, the leading banking institution in Perú, managing over 1,790 ATMs and 340 branches. This end-to-end solution significantly enhanced operational efficiency, improving ATM uptime by 4.2% and reducing cash logistics costs by 8.7% while ensuring sufficient and reliable provisioning across all locations. The system integrates 42 machine learning models to forecast daily demand across three channels and two currencies (local and US\$) which feed into a linear optimization framework that incorporates complex business constraints. We also discuss new modules for emerging ATM types, and extending the optimization framework to include positive interest cash-flows that emerge from temporal cash deposits at the Central Bank of Perú.

2 - Privacy of the Last Iterate in Cyclically-Traversed DP-Sgd on Nonconvex Composite Losses

Weiwei Kong, Google Research, New York, NY, United States, Monica Ribero

Differentially-private stochastic gradient descent (DP-SGD) is a family of iterative machine learning training algorithms that privatize gradients to generate a sequence of differentially-private (DP) model parameters. It is also the standard tool used to train DP models in practice, even though most users are only interested in protecting the privacy of the final model. Tight DP accounting for the last iterate would minimize the amount of noise required while maintaining the same privacy guarantee and potentially increasing model utility. However, last-iterate accounting is challenging, and existing works require strong assumptions not satisfied by most implementations. These include assuming (i) the global sensitivity constant is known - to avoid gradient clipping; (ii) the loss function is Lipschitz or convex; and (iii) input batches are sampled randomly.

In this work, we forego any unrealistic assumptions and provide privacy bounds for the most commonly used variant of DP-SGD, in which data is traversed cyclically, gradients are clipped, and only the last model is released. More specifically, we establish new Renyi differential privacy (RDP) upper bounds for the last iterate under realistic assumptions of small stepsize and Lipschitz smoothness of the loss function. Our general bounds also recover the special-case convex bounds when the weak-convexity parameter of the objective function approaches zero and no clipping is performed. The approach itself leverages optimal transport techniques for last iterate bounds, which is a nontrivial task when the data is traversed cyclically and the loss function is nonconvex.

3 - Finding the Best Arm Privately in Unimodal Bandits

Irem Nur Keskin, Duke University, Durham, NC, United States, Jiaming Xu

This paper introduces a novel private bandit learning problem where the learner aims to find the highest reward arm in a continuous set with unknown, unimodal expected rewards. The learner, facing an adversary who observes their actions but not the rewards, must minimize arm queries while preserving privacy. We propose the Replicated Golden Section Search (RGSS) strategy, which efficiently narrows down the subset of potentially optimal arms while replicating arm queries to obscure adversary observations. For an α -smooth unimodal reward function and a target privacy level L , our analysis demonstrates that RGSS can locate the best arm within ϵ -accuracy using $O(L \log(1/\epsilon)/\epsilon^{2\alpha})$ queries, while hindering the adversary's success to a probability no greater than $1/L$. Additionally, we introduce an adaptive algorithm that eliminates the need for prior knowledge of α . We further establish a lower bound, showing that any private strategy requires at least $\Omega(L/\epsilon^{2\alpha})$ arm queries. These findings reveal that our methods are nearly optimal, with the multiplicative factor of L being a necessary price to pay for privacy. This contrasts sharply with prior research in first-order convex optimization, which suggested that an additive L factor was sufficient for privacy preservation.

4 - Nonlinear Fixed Points and Stationarity: Economic Applications

Giacomo Lanzani, Harvard University, Cambridge, MA, United States, Simone Cerreia-Vioglio, Roberto Corrao

We consider the fixed points of nonlinear operators that naturally arise in games and general equilibrium models with endogenous networks, in dynamic programming, in models of opinion dynamics with stubborn agents, and in financial networks. We study limit cases that correspond to high coordination motives, infinite patience, vanishing stubbornness, and small exposure to the real sector in the applications above. Under monotonicity and continuity assumptions, we provide explicit expressions for the limit fixed points. We show that, under differentiability, the limit fixed point is linear in the initial conditions and characterized by the Jacobian of the operator at any constant vector with an explicit and linear rate of convergence. Without differentiability, but under additional concavity properties, the multiplicity of Jacobians is resolved by a representation of the limit fixed point as a maxmin functional evaluated at the initial conditions. In our applications, we use these results to characterize the limit equilibrium actions, prices, and endogenous networks, show the existence and give the formula of the asymptotic value in a class of zero-sum stochastic games with a continuum of actions, compute a nonlinear version of the eigenvector centrality of agents in networks, and characterize the equilibrium loss evaluations in financial networks.

5 - Presolving Techniques for Minlps

Anna Deza, University of California-Berkeley, Berkeley, CA, United States, Andres Gomez, Alper Atamturk

We present pre-processing techniques for scaling mixed integer non-linear optimization problems with indicator variables, including mixed-integer conic and semidefinite programs. A growing body of research has produced strong relaxations for many such problems, often yielding very tight bounds. However, in contrast to mixed-integer linear or quadratic settings, current commercial solvers often fail to fully exploit these relaxations during branch-and-bound operations, resulting in limited scalability. In this work, we show how to leverage strong relaxations in order to infer optimality conditions, fix variables, and add cuts prior to the full optimization step. To illustrate the significant potential for speed-up and enhanced scalability, we apply our techniques to mixed-integer quadratic multi-period problems and sparse principle-component analysis.

TE48

Building B Level 3 B308

Advanced Analytics for Decision Support

Invited Session

Decision Analysis Society

Chair: Hyungjoo Cha, Korea University, Seoul, Korea, Republic of

1 - Brainnormalizer: Reconstructing Healthy Brain Images from Tumor Mris Using Masked Controlnet with Edge Conditioning

Mingu Kwak, Georgia Institute of Technology, Atlanta, GA, United States, Yeonju Lee, Hairong Wang, Jing Li

Brain tumors significantly alter neural tissue and cause global deformation, creating challenges for clinicians attempting to visualize tumor boundaries and progression. Current methods often struggle to accurately reconstruct healthy brain representations from tumor-affected images, limiting intuitive comparative analysis. We propose BrainNormalizer, a novel approach using ControlNet diffusion models with three key contributions: (1) a specialized masking strategy during fine-tuning that focuses on tumor region restoration, (2) edge map conditioning that addresses global deformation while preserving brain structure, and (3) strategic text prompts that instruct the model to interpret tumor images as healthy ones. BrainNormalizer effectively transforms tumor-containing images into their healthy counterparts by conditioning the generative process with the carefully designed inputs. Evaluations on the Brain Tumor Segmentation (BraTS) dataset demonstrate that our approach successfully generates high-quality healthy brain images while maintaining structural integrity. BrainNormalizer outperforms existing methods particularly in handling complex deformation patterns and preserving anatomical consistency. This work provides clinicians with an intuitive comparative visualization tool that enhances diagnostic capabilities and offers potential applications in treatment planning, medical education, and supplementary analysis for neuro-oncological tasks.

2 - Quantum Boolean Circuit Design via Gate Pattern-Based Column Generation

Jihye Jung, Georgia Institute of Technology, Atlanta, GA, United States, Jae Hyeok Lee, Changhyun Kwon, Jaehwan Lee

Quantum computing has recently gained attention for the potential boost it promises to computational power by exploiting the probabilistic nature of quantum mechanical concepts. That being said, development on quantum machines has taken major steps so far. However, a considerable gap still exists in building a fully reliable quantum machine, highlighting the strong need for error minimization. One thread of these efforts is to design an efficient quantum circuit with the fewest operational units required to implement the target computation, which leads to a complex decision-making process. This study specifically focuses on implementing Boolean functions in the quantum environment using a series of logical gates called Multiple-Control Toffoli gates, where the exponential space of gate configurations directly determines how the gate operates. Few studies have already proposed optimization-based approaches to this problem, but they still face scalability challenges due to the exponential nature inherent in quantum computation. We explore the capability of our new mixed-integer programming model that formulates the problem directly based on gate configuration patterns.

3 - Centralized Control for Integrated Production and Distribution Planning via Reinforcement Learning

Kim Haneul, Korea University, Seoul, Korea, Republic of, Taesu Cheong

Deep reinforcement learning (DRL) has recently emerged as a promising approach for addressing multi-echelon supply chain management problems under stochastic environments. However, existing DRL-based studies often fall short in capturing the operational constraints and structural intricacies of real-world supply chains. Moreover, there remains a significant lack of empirical evidence validating the transferability of learned policies across varying demand distributions. In this study, we propose a DRL model that makes node-level production and edge-level distribution decisions. The policy network employs a spatio-temporal graph neural network (GNN) encoder to capture supply chain dynamics. Furthermore, the model incorporates structured policy design through action chunking, enabling it to effectively accommodate various operational constraints. Experiments are conducted to compare the proposed method with constrained reinforcement learning baselines in terms of learning efficiency. Robustness against alternative policies is validated, and transferability is assessed across a range of demand variability scenarios.

4 - An Exact Algorithm for Vehicle Routing Problem with Availability Profiles

Hyungjoo Cha, Korea University, Seoul, Korea, Republic of, Joonghoo Park, Chungmok Lee, Taesu Cheong

This talk presents a branch-and-cut approach for the VRP with Availability Profiles (VRPAP). The VRPAP generalizes the well-known VRPTW by introducing multiple time windows with variable values available upon arrival. We present a novel extended formulation for the problem that provides a stronger LP relaxation bound than the previous one in the literature. We then develop a branch-and-cut algorithm to address the exponentially many constraints efficiently. Moreover, we also show that the optimal AP value on the fixed routes can be obtained by dynamic programming, which is incorporated into the branch-and-cut algorithm as a primal heuristic. Computational experiments show that the proposed algorithm outperforms the existing approaches.

5 - Uncertainty Aware Tweedie Hybrid Loss for Intermittent and Erratic Demand Forecasting

Hwayong Jeong, Kwangwoon University, Nowon-gu, Korea, Republic of, Seungjoon Baek, Jiyun Woo, Younggi Kim, Haejoong Kim, Dae-Eun Lim, Sangmin Lee

Online retail demand often exhibits intermittent zero-inflation and erratic spikes. These violate normality and homoscedasticity assumptions, challenging variance-sensitive loss functions like MSE. We present a hybrid loss which combines Tweedie negative log-likelihood and Huber losses for forecasting irregular demand patterns. Further, we enhance forecasting robustness by reweighting loss via aleatoric and adjusting learning rates via epistemic uncertainty. Our method is implemented on N-Beats to jointly predict the mean and dispersion of a Tweedie distribution. Experiments show 13.27% lower MAE, 9.31% lower RMSE, and 5.19% lower MAPE compared to baseline methods. The results confirm that the proposed approach effectively handles zero-inflation, outliers, and instability in demand, leading to better predictions.

TE49

Building B Level 3 B309

Data-Driven Digital Commerce

Invited Session

MSOM:Technology, Innovation, and Entrepreneurship

Chair: Ruomeng Cui, Emory University, Amazon, Decatur, GA, United States

Co-Chair: Peibo Zhang, Emory University, Atlanta, United States

1 - Reputation Farming: A Structural Model of Freelancers' Bidding Strategy

Xi Nan, University of Washington, Seattle, WA, United States, Xingchen Xu, Yong Tan

The online freelance labor market has experienced rapid growth driven by advances in digital technologies and a rising demand for flexible work arrangements. To mitigate information asymmetries, most platforms implement reputation systems that enable clients to provide reviews, thereby guiding the choices of future clients. However, this structure imposes a significant entry barrier for new freelancers, who encounter a “cold start” problem: without initial reviews, they struggle to secure jobs, and without jobs, they cannot obtain reviews or build their reputation. This dynamic may incentivize freelancers to bid strategically, submitting lower bids early in their tenure and gradually increasing bids as their reputational capital grows. To examine the prevalence of such strategic behavior, we develop a dynamic discrete choice model with multi-level heterogeneities and empirically estimate it using a large-scale dataset. Our findings reveal that over 20% of freelancers exhibit forward-looking behavior, with an approximately 5% higher likelihood of bidding lower-tier prices compared to their myopic counterfactuals. Additional robustness checks and policy simulations are conducted to offer market design recommendations for improving efficiency and welfare.

2 - Estimating Effects of Long-Term Treatments

Jinglong Zhao, Boston University, Boston, MA, United States, Shan Huang, Chen Wang, Yuan Yuan, Brocco Zhang

A/B tests struggle to capture long-term treatment effects, as product changes last far beyond typical experiments. We propose a longitudinal surrogate framework that links long-term outcomes to user attributes, short-term metrics, and treatments. Tested on two million-user WeChat experiments, our approach outperforms existing methods.

3 - Learning How to Learn: Adaptive Algorithms for Long-Term Memory

Xufei Liu, Wharton, University of Pennsylvania, Philadelphia, PA, United States, Kenneth Moon, Gad Allon

Algorithms present tremendous potential for guiding human learning in many contexts, from markets to workplaces. Yet, there exists few data-driven methods which actively recognize and exploit the relevant structure of human knowledge. We collaborate with a popular language app to design dynamic algorithms that achieve lasting learning by teaching users to gradually strengthen their use of linguistic context as a form of broader knowledge, in order to drive more efficient and durable learning outcomes. Our work has three primary goals: (1) to model how knowledge evolves during language learning and identify key drivers of accelerated learning; (2) to capture

how knowledge transfers between related tasks; and (3) to optimally schedule review intervals that both maximize retention and enhance students' long-term learning capabilities. We use a student's latent memory parameter distributions to descriptively find similarities and patterns for students of the same caliber. Then, we can construct a dynamic programming question that will maximize their reward by shifting their distribution of learning parameters so that students learn faster and more effectively over time.

4 - Ship-Then-Shop as a Promotional Tool: Evidence from An Online Omnichannel Fashion Retailer

Yiwei Wang, Zhejiang University, Haining, China, People's Republic of

The ship-then-shop (STS) model allows customers to receive a preselected assortment of products to try at home before making purchase decisions. While STS enhances the online shopping experience by providing access to physical products, its impact on retailer profitability remains largely unexplored. Using field experiment data from an online omnichannel fashion retailer, we study the impact of a short-term non-subscription STS promotional program on customer behavior. We uncover a demand spillover effect of STS promotion: beyond their direct purchases from STS shipments, treated customers exhibited increased transactions in the retailer's online sales and subscription rental channels. During the treatment period, demand spillovers to online sales were weaker than those to subscription rentals. After the program ended, demand spillovers to online sales amplified, while those to subscription rentals declined. We provide suggestive evidence for four mechanisms underlying these patterns. First, STS shipments increased sales of both promoted (i.e., included) brands and unpromoted brands—what we refer to as the brand promotion and customer engagement effects, respectively. However, both effects were limited to previously consumed brands. Second, ending the STS program led to a demand migration to online sales of unpromoted, previously consumed brands, suggesting that the brand promotion effect was transient, while the customer engagement effect was persistent. Third, STS shipments exerted a reminder effect on rental subscription customers. Fourth, customers exhibited cross-channel complementarity between buying and renting. Based on the heterogeneous profit implications of STS promotion across customer segments, we propose a data-driven customer targeting strategy.

5 - When Clinicians Edit AI: Classifying and Measuring Documentation Errors in Primary Care

Ruozhu (Savannah) Wang, Boston university, Boston, MA, United States

AI-based ambient documentation tools that automatically transcribe clinician-patient conversations have been increasingly used to reduce the burden in primary care. However, there are limited studies investigating the documentation errors made by AI and clinicians' ability to manage them. Using a dataset of clinician-patient consultation vignettes and corresponding medical notes, we identify and categorize previously unrecognized documentation errors made by clinicians and AI. We develop two taxonomies to classify the error types present in clinician-written and AI-generated notes, and investigate clinicians' performance by quantifying their ability to identify and correct documentation errors in clinician-written and AI-generated notes. This study provides a novel approach for assessing clinician-AI collaboration and contributes to the broader understanding of care quality in AI-assisted clinical environments.

TE50

Building B Level 3 B310

AI and Operations

Invited Session

MSOM: Service Operations

Chair: Hongseok Namkoong, Columbia University, New York, NY, United States

1 - Active Learning: Recent Developments, Current State, and Open Challenges

Steve Musmann, Georgia Institute of Technology, Atlanta, GA, United States

Active learning is a vital component in the machine learning toolbox, particularly in scenarios where labeled data is scarce or expensive to obtain. This talk surveys developments in active learning, including recent innovations in query strategies, novel use cases, and the impact of advances in representation learning. The current landscape will be contextualized through empirical findings, conceptual insights, and case studies. This talk aims to provide a critical synthesis of the field and outline promising directions for future research in active learning within the broader machine learning ecosystem.

2 - Paper Skygest: Personalized and Personalizable Recommendations for Academics on Bluesky

Sophie Greenwood, Cornell Tech, New York, NY, United States, Nikhil Garg

Social media platforms are an important way for many academics to announce emergent work and discuss existing work with the global research community. Which users see a post is intermediated by the feed recommendation and ranking algorithm. On most platforms, the recommendation algorithm is (1) opaque and outside a platform user's direct control and (2) not built with academics in mind. A large cohort of researchers recently migrated to Bluesky, a rising social media platform that allows individuals to build feeds for themselves and other users. We developed a personalized feed for academic content on Bluesky, *Paper Skygest*, which shows a user posts about papers by accounts she is following, in reverse chronological order. We launched publicly in March 2025; since then the feed has had substantial usage. In this presentation, we will give an introduction to Bluesky's decentralized feed ecosystem and describe ongoing work leveraging Paper Skygest to study academic recommendations. In particular, we aim to explore users' preferences for fine-grained control over their recommendations, as well as the market-level implications of user control.

3 - Impatient Bandits: Optimizing for the Long-Term Without Delay

Kelly Zhang, Imperial College London, London, United Kingdom, Thomas Baldwin-McDonald, Kamil Ciosek, Lucas Maystre, Daniel Russo

Increasingly, recommender systems are tasked with improving users' long-term satisfaction. In this context, we study a content exploration task, which we formalize as a bandit problem with delayed rewards. There is an apparent trade-off in choosing the learning signal: waiting for the full reward to become available might take several weeks, slowing the rate of learning, whereas using short-term proxy rewards reflects the actual long-term goal only imperfectly. First, we develop a predictive model of delayed rewards that incorporates all information obtained to date. Rewards as well as shorter-term surrogate outcomes are combined through a Bayesian filter to obtain a probabilistic belief. Second, we devise a bandit algorithm that quickly learns to identify content aligned with long-term success using this new predictive model. We prove a regret bound for our algorithm that depends on the $\text{Value of Progressive Feedback}$, an information theoretic metric that captures the quality of short-term leading indicators that are observed prior to the long-term reward. We apply our approach to a podcast recommendation problem, where we seek to recommend shows that users engage with repeatedly over two months. We empirically validate that our approach significantly outperforms methods that optimize for short-term proxies or rely solely on delayed rewards, as demonstrated by an A/B test in a recommendation system that serves hundreds of millions of users.

4 - Advancing Synthetic Control Methods for Individual Level Causal Inference

Saeyoung Rho, Columbia University, New York, NY, United States

Synthetic control is a widely used method in causal inference for evaluating policy interventions, such as tariffs or tax reforms. While traditionally applied to aggregate-level data, recent work has begun to extend synthetic control to disaggregated datasets, including individual health records and targeted marketing data. Such datasets pose new challenges: we may observe vastly more units (e.g., millions of individuals instead of a single state) or obtain much longer time series for a smaller cohort (e.g., longitudinal studies). These distributional shifts call into question the assumptions underlying standard synthetic control methods. In this talk, I will present two new algorithms that adapt synthetic control to individual-level analyses. Cluster Synthetic Control (ClusterSC) is designed for settings with many units but short time horizons; it introduces a donor selection step to improve the accuracy of prediction. Time-Aware Synthetic Control (TASC) addresses settings with few units but long time series, leveraging temporal dynamics in a latent space to better estimate counterfactuals. Both methods are theoretically grounded and achieve provable accuracy gains across distinct data regimes, offering new tools for extending synthetic control to richer and more complex datasets.

TE51

Building B Level 3 B311

Analytics and Aisles: Bridging Theory and Practice in Retail Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Rim Hariss, McGill University, Montreal, QC, Canada

1 - Out-of-Stock, Out-of-Store? Estimating Cross-Store Substitution via Customer Journey Tracking

Rim Hariss, McGill University, Montreal, QC, Canada, Sanjith Gopalakrishnan, Hamidreza Moshrefi, Saibal Ray

Stockouts remain a persistent and costly challenge in fashion retail, where limited replenishment and evolving customer preferences amplify lost sales risk. In this work, we leverage a novel dataset from a leading footwear retailer, combining transaction records, daily inventory, and real-time session data from an in-store tracking tool. This tool logs customer product requests and associate interactions, providing a clickstream-like view of the offline shopping journey.

We use these data to: (1) estimate the rate at which customers accept cross-store fulfillment when their desired product is unavailable; (2) identify the product, customer, and store-level factors that drive substitution decisions; (3) quantify the incremental value of enabling cross-store substitutions on sales and margins; and (4) explore how substitution patterns can inform optimal inventory allocation across stores. Our analysis reveals significant heterogeneity in substitution likelihood and demonstrates that incorporating these patterns into allocation decisions can substantially improve performance.

Our findings offer new insights into consumer behavior under stockouts and provide actionable guidance for inventory and assortment planning in fashion retail, bridging the gap between online and offline behavioral data.

2 - Optimal Replenishment Policies for Industrial Vending Machines

Esma Gel, University of Nebraska-Lincoln, Lincoln, NE, United States, Karina Sindermann, Nesim Erkip

Our paper introduces and analyzes optimal replenishment strategies for Industrial Vending Machines (IVMs) used by industrial supply distributors managing customer's inventory of maintenance, repair and operational supplies. IVMs have emerged as an important technology, offering real-time inventory tracking and enhanced supply chain visibility in vendor-managed inventory settings. We formulate this novel replenishment control problem where multiple items with stochastic demand face a joint fixed replenishment cost and item-specific stockout costs. We develop two approaches: (1) a state-dependent trigger set control policy and (2) a fixed cycle replenishment policy. For the trigger set policy, we address the previously intractable challenge of determining which inventory states should optimally trigger replenishment by reformulating the problem as an optimal stopping rule and proving key monotonicity properties. This innovation enables us to characterize the optimal policy and develop computationally efficient approximations that achieve near-optimal performance. For fixed cycle replenishment, we derive closed-form expressions for the optimal cycle length. Our numerical studies demonstrate that our approximate online trigger set policy remains within 1% of optimality while computing thousands of times faster. A case study using real transaction data from 316 item-device combinations shows that optimal fixed cycle replenishments could reduce costs by 61.7 to 78.6% compared to current daily replenishment practices, while the trigger set control policy provides additional savings of 4.1 to 22.9%. These results provide industrial distributors with theoretically sound, practical tools for optimizing IVM operations.

3 - Data Driven Assortment Optimization with Product Returns

Recep Yusuf Bekci, University of Waterloo, Waterloo, ON, Canada, Mehmet Gumus, Sentao Miao

The rise of online shopping and lenient return policies has led to a significant increase in product returns, posing economic and environmental challenges. In collaboration with a European fast fashion retailer, this study addresses how to leverage returns data in product assortment decisions. We tackle two main challenges: dynamically learning customer preferences and estimating return rates amid delayed returns. We formulate an assortment optimization problem considering customer and return uncertainties and propose an algorithm to optimize decisions. Our performance guarantee and analyses with synthetic and real data provide insights for retailers on managing returns and improving assortment decisions.

4 - Distributionally Robust Multi-Purchase Choice Modeling and Assortment Optimization

Joline Uichanco, University of Michigan, Ross School of Business, Ann Arbor, MI, United States, Jiawei Zhang, Zhuoxin Chen

In many retail environments, consumers often purchase multiple products in a single transaction, yet most existing choice models focus solely on single-purchase behavior. This talk introduces a new *Marginal Distribution Multi-Purchase* (MDM-MP) choice model that extends traditional choice frameworks by accommodating multiple purchases under uncertainty in random utility correlations. Leveraging a distributionally robust optimization approach, our model assumes known marginal distributions of random utility while remaining agnostic about their joint distribution.

We derive tractable formulations for choice probabilities under this framework and explore their implications for assortment optimization. Notably, we show that the optimal assortments under the MDM-MP model are not necessarily revenue-ordered, challenging a key result from the multinomial logit (MNL) model literature. We also provide closed-form solutions under exponential noise assumptions, enabling efficient implementation and estimation.

Our results yield actionable insights for retailers designing assortments in contexts where customers often buy more than one item—such as flash sales or e-commerce platforms—highlighting the strategic value of "anchor" products that consistently drive multi-item purchases. This work contributes to the growing literature on robust and multi-choice consumer models with practical relevance for revenue management and operations strategy.

TE52

Building B Level 3 B312

Empirical and Analytical Research on Sustainable Operations

Invited Session

MSOM: Sustainable Operations

Chair: Mirel Yavuz, IESE Business School, BARCELONA, Spain

Co-Chair: Anna Saez de Tejada Cuenca, IESE Business School, BARCELONA, 08034, Spain

1 - Interactive Preference-Based Optimization with Unknown Value Function: Application to Sustainable Sourcing In the Apparel Supply Chain

Mirel Yavuz, IESE Business School, BARCELONA, Spain, Charles Corbett

Optimization problems often inherently include multiple criteria, and the underlying value function of the decision maker is often unknown and difficult to elicit. This is particularly true in sustainable operations, when some of the criteria are environmental or social in nature, and frequently in conflict with one another. Firms seeking to incorporate sustainability in their decisions face difficult choices, with no well-defined guidelines on how to make trade-offs between different environmental and social impact categories. In this work, we consider a multi-objective linear optimization problem, and assume the preferences of the decision maker are based on an unknown implicit linear value function. We propose an interactive optimization approach that asks pairwise comparison questions to the decision maker to determine the solution that is most aligned with her preferences. The method is guaranteed to yield the decision maker's optimal solution, without needing to elicit the weights defining their value function. To illustrate the method, we use the context of sustainable sourcing in the apparel industry. We numerically test our approach using realistic data based on the Higg Material Sustainability Index (MSI) with three to four criteria: cost, global warming potential, water use and fossil fuel use. Preliminary interactions with

practitioners suggest they find this approach promising, and that it can help them better understand and articulate their preferences. We hope that this work will draw attention to this type of interactive method for solving multi-objective optimization problems with unknown or implicit value functions in sustainable operations.

2 - Is (Climate) Adversity the Mother of Invention? Evidence from Billion Dollar Weather Disasters

Suvrat Dhanorkar, The Pennsylvania State University, University Park, PA, United States, Suresh Muthulingam, Max Ji

We examine the impact of extreme weather events on firm-level innovation.

3 - An Empirical Investigation of Establishment-Level Operations and Occupational Hazards in Underserved Communities

Abhinav Shubham, Georgia Institute of Technology, Atlanta, GA, United States, Ravi Subramanian

Socioeconomic factors and differing establishment-level operational characteristics may contribute to disparities in exposure to occupational safety hazards. We investigate whether similar establishments pose disparate occupational injury risks when operating in underserved communities. Additionally, we consider the moderating role of random inspections by regulators and successful unionization efforts by employees in mitigating disparities in occupational injury risks.

4 - Specialize OR Balance? Sales Agent Allocation in the Presence of Low-Resilience Firms at the Base of the Pyramid

Olumurejiwa Fatunde, Georgetown University, Washington, DC, United States, Gonzalo Romero, Andre Calmon, Joann de Zegher

This paper explores whether distributors should assign some sales agents to “specialize” in serving informal firms in settings where such firms are particularly vulnerable. We propose a novel hypothesis for why distributors struggle to scale last-mile distribution of durable, life-improving goods to customers at the Base of the Pyramid (BoP): the relatively lower supply chain relationship resilience of informal retailers, which are the primary source of retail purchases for BoP customers. We measure this concept as the ratio of post-disruption revenue to pre-disruption revenue. Informal retailers are often embedded within communities where local, long-term relationships are essential to business transactions. We build an analytical model to demonstrate how supply chain relationship resilience can shape a distributor’s allocation of retailers to sales agents. We show that if resilience levels are uniform across informal and formal retailers, the distributor’s optimal assignment to sales agents is fully unbalanced (i.e., the majority of sales agents handle only one of the two retailer types). We then show how accounting for differences in supply chain relationship resilience can dramatically change the firm’s optimal sales agent assignment strategy. We provide empirical evidence for this hypothesis by analyzing data from a distributor selling to 319 formal retailers and 505 informal retailers in India from April 2016 to December 2019. Our results demonstrate the importance of supply chain relationship resilience when designing BoP operations. In particular, informal retailers’ resilience levels should be considered when making resource allocation decisions and designing BoP operations strategy.

TE53

Building B Level 3 B313

Frontier Research on Revenue Management and Pricing

Invited Session

Revenue Management

Chair: Zhoupeng (Jack) Zhang, University at Buffalo School of Management, The State University of New York, Buffalo, NY, United States

Co-Chair: Titing Cui

1 - Employees OR on-Demand Workers? Rethinking Workforce Strategy in a Productivity-Driven Tech Era

Tolga Dizdärer, Boston College, Carroll School of Management, Chestnut Hill, MA, United States, Dmitry Mitrofanov

This paper examines optimal workforce design strategies in environments where firms can choose between permanent employees and on demand independent contractors. We consider a setting where the same workload can be distributed across different types of workers, each with distinct cost structures. Although both worker types can perform similar tasks, their effectiveness varies in subtle but important ways. It is not obvious how firms should determine the optimal mix of workers—or how recent advances in automation and AI may shift this balance. We show that as worker productivity changes, both the optimal staffing levels and the composition of the workforce can shift dramatically. In particular, we characterize conditions under which firms rely entirely on one worker type, split work across both, or adjust their strategy in response to workload intensity. The results offer insights for workforce design in service operations, platform businesses, and environments facing uncertainty due to technological change.

2 - Optimal Design for Revenue-Based Financing

Eryn Juan He, University of Utah, Salt Lake City, UT, United States, Joel Goh

Small businesses play a crucial role in emerging markets, despite facing various challenges that impede their growth and success. One primary challenge is the limited access to affordable financing. Revenue-based financing (RBF) has emerged as an alternative, enabling small businesses to secure funding without sacrificing part of its equity (compared with equity-based financing) or the burden of inflexible debt conditions (compared with debt-based financing). Under RBF, repayments are structured as a percentage of future revenues, aligning the interests of lenders and borrowers.

In this paper, we develop a stochastic model to examine the conditions under which a lending platform can achieve higher profitability through revenue-based financing compared to classical debt financing. Our analysis yields two key insights. First, we identify the conditions that make RBF advantageous for platforms, highlighting the roles of revenue mean and volatility, as well as the variability of operating costs. Second, building on this analysis, we design an optimal contract framework that allows platforms to effectively screen different types of retailers, thereby maximizing their expected utility.

3 - Competing with Genai: a Two-Period Game-Theoretic Model

Guoxing He, Faculty of Business and Economics, HKU, Hong Kong, Hong Kong, Chunghan Kang, Feng Tian

Two heterogeneous firms compete using AI in high- and low-margin markets over two periods. First-period AI usage influences second-period productivity gains, with spillover effects on the competitor. Depending on the spillover rate and initial lead in the high-margin market, these spillover effects can undermine the leading firm’s competitive advantage.

TE54

Building B Level 2 B216

Recent Advances in Online Resource Allocation Problems

Invited Session

Revenue Management

Chair: Murray Lei, Queen's University, Kingston, ON, Canada

1 - Privacy Preserving Dynamic Assortment Selection

Young Hyun Cho, Purdue University, West Lafayette, IN, United States, Will Wei Sun

This presentation introduces a novel framework for privacy-preserving dynamic assortment selection using the multinomial logit (MNL) bandits model. Our approach employs a perturbed upper confidence bound method that satisfies Joint Differential Privacy (JDP), a robust privacy guarantee well-suited for dynamic environments. We will present the theoretical analysis of our framework, which includes a near-optimal regret bound that quantifies the trade-off between privacy and performance. Finally, we demonstrate the superiority of our method over the benchmark through extensive simulations and a case study on the Expedia hotel dataset.

2 - Privacy-Aware Online Resource Allocation with Privacy Preference Heterogeneity

Geoffrey Chua, Nanyang Technological University, Singapore, Singapore, Du Chen, Frank Chen

Online resource allocation is a central problem in revenue management that has been perplexed recently by growing privacy concerns, as decisions often rely on customers' sensitive data. The question of how to efficiently allocate resources while protecting customer privacy has become imperative, yet remains largely unexplored. In this paper, we study online allocation problems with privacy considerations, where customers arrive sequentially, each submitting a service request containing sensitive data along with a privacy request specifying a desired protection level. For each customer, after receiving both requests, the decision maker must determine an allocation immediately and irrevocably, while meeting the requested privacy level under the joint differential privacy framework. The objective is to maximize total revenue subject to limited resources. To address this, we develop a generalized algorithm based on the primal-dual paradigm, incorporating a noisy shadow price to coordinate allocations. The algorithm employs a tree-based mechanism for privately tracking resource consumption and uses mirror descent to update shadow prices. The tree-based mechanism is further customized to accommodate heterogeneous privacy preferences. We prove that the algorithm achieves sublinear regret and satisfies the requested privacy guarantees across a broad class of settings, including both stationary and nonstationary arrivals. Moreover, we identify a spillover effect in which protecting one customer's privacy may inadvertently degrade the utility for others. We show that collecting privacy preferences in advance and adjusting the algorithm accordingly can substantially mitigate this negative impact. Numerical experiments confirm our theoretical findings and demonstrate the effectiveness of the proposed algorithm.

3 - Near-Optimal Pricing and Resource Allocation in a Large-Scale Service System

Sun Xu, Miami Herbert Business School, Coral Gables, FL, United States

We study dynamic pricing and capacity allocation in large-scale service systems where multiple service units serve customers who are both price- and delay-sensitive. Customers are segmented into classes, each characterized by class-specific service rates and demand functions shaped by posted prices and perceived delays. To jointly optimize revenue and delay performance, we propose a family of state-dependent greedy heuristics. These heuristics (i) assign dedicated service capacities to each customer class and (ii) dynamically set prices by solving a tractable one-step optimization problem at each decision epoch. Despite their simplicity, the proposed policies achieve provably strong performance: in large-market regimes where demand scales with the number of servers n , they attain a relative optimality gap of $O(n^{-(2/3)} \log n)$. Moreover, we show that this performance guarantee extends to any work-conserving allocation, not just the dedicated-capacity rule. This robustness highlights the practicality of our approach in settings where centralized control over service units is limited. Extensive numerical experiments validate the theoretical results and offer further managerial insights. In particular, we observe that queue-based wait-time estimates can lead to non-monotonic congestion pricing behavior—i.e., optimal prices may decrease with increasing congestion—a counterintuitive yet rational response under certain demand sensitivities.

4 - Diffusion Approximation in Restless Bandits

Chen YAN, University of Michigan, Ann Arbor, MI, United States

We study finite-horizon restless multi-armed bandits (RMAB) with homogeneous arms, focusing on degenerate cases where LP-based policies suffer a suboptimal gap. We introduce a novel diffusion approximation that captures both mean and variance, achieving significantly tighter performance guarantees and showing that standard fluid approximations are loose in degenerate settings.

5 - Capacitated Multi-Echelon Batch Inventory Networks

Maximiliano Stock, The University of Chicago Booth School of Business, Chicago, IL, United States, Levi DeValve

We study the Capacitated Multi-echelon Batch Inventory problem, which models batch inventory ordering in a two-echelon fulfillment network with capacity constraints at the lower echelon, as often seen in e-retailing. A centralized upper echelon places batch orders from a supplier and allocates units to the lower echelon to minimize total costs, including ordering, allocation, and expected fulfillment costs.

The hard capacity constraints make the problem structurally similar to capacitated facility location, which is well known to be challenging to approximate using linear relaxations. We address these challenges by exploiting key features of our setting: the allocation and fulfillment costs are metric, and the upper echelon is uncapacitated.

Leveraging these features, we develop a new clustering technique that uses information from the dual program to group nearby distribution centers in the lower echelon with enough inventory. This clustering enables us to round the LP solution while satisfying the capacity constraints and leads to a constant-factor approximation algorithm for the offline two-stage version of the problem.

We further extend our approach to an online setting in which demand arrives sequentially over a finite time horizon. To address this, we develop a framework that integrates our approximation algorithm with tools from the online decision-making literature to decide on the ordering, allocation, and online fulfillment. Our framework achieves an additive regret bound that decreases with the time horizon, which makes the theoretical guarantee of the online solution to converge to that of the offline two-stage problem as the horizon grows.

TE55

Building B Level 3 B314

Online Decision Making Algorithms

Invited Session

Revenue Management

Chair: Jinglong Zhao, Boston University, Boston, MA, United States

Co-Chair: Mika Sumida, University of Southern California, Los Angeles, CA, United States

1 - Randomized Policy Optimization for Optimal Stopping

Xinyi Guan, The Hong Kong Polytechnic University, Hong Kong, China, People's Republic of, Velibor Misic

Optimal stopping is the problem of determining when to stop a stochastic system in order to maximize reward, which is of practical importance in domains such as finance, operations management and healthcare. Existing methods for high-dimensional optimal stopping that are popular in practice produce deterministic linear policies – policies that deterministically stop based on the sign of a weighted sum of basis functions – but are not guaranteed to find the optimal policy within this policy class given a fixed basis function architecture. In this paper, we propose a new methodology for optimal stopping based on randomized linear policies, which choose to stop with a probability that is determined by a weighted sum of basis functions. We motivate these policies by establishing that under mild conditions, given a fixed basis function architecture, optimizing over randomized linear policies is equivalent to optimizing over deterministic linear policies. We formulate the problem of learning randomized linear policies from data as a smooth non-convex sample average approximation (SAA) problem. We theoretically prove the almost sure convergence of our randomized policy SAA problem and establish bounds on the out-of-sample performance of randomized policies obtained from our SAA problem based on Rademacher complexity. We also show that the SAA problem is in general NP-Hard, and consequently develop a practical heuristic for solving our randomized policy problem. Through numerical experiments on a benchmark family of option pricing problem instances, we show that our approach can substantially outperform state-of-the-art methods.

2 - Incentive-Aware Dynamic Resource Allocation Under Long-Term Cost Constraints

Yan Dai, Massachusetts Institute of Technology, Cambridge, MA, United States, Negin Golrezaei, Patrick Jaillet

Motivated by public and societal challenges such as governments allocating mobile health units across underserved regions, we study the dynamic allocation of a reusable public resource to self-interested agents -- who may strategize based on their private valuations for the resource -- under long-term cost constraints. Our objective is to simultaneously (i) maximize social welfare, i.e., yielding efficient utilization of the public resource, (ii) satisfy multi-dimensional long-term cost constraints, and (iii)

incentivize truthful reporting. We begin by numerically evaluating primal-dual methods widely used in non-strategic online resource allocation, and find them to be highly fragile in strategic settings. Indeed, agents easily manipulate their reports to distort future dual updates for future gain.

To address this vulnerability, we develop an incentive-aware framework that makes primal-dual methods robust to strategic behavior. Our design combines epoch-based lazy updates -- where dual variables remain fixed within each epoch -- with randomized exploration rounds that extract approximately truthful signals for learning. Leveraging carefully designed novel online learning techniques that can be of independent interest for dual updates, our mechanism achieves $\sim O(\sqrt{T})$ social welfare regret, satisfies all cost constraints, and ensures incentive alignment. This matches the performance of non-strategic allocation approaches while being robust to strategic agents.

3 - On the Sample Bias of Bandit Algorithms

Yilun Chen, the Chinese University of Hong Kong (Shenzhen) , Shenzhen, China, People's Republic of, Jiaqi Lu

Many real-world datasets are collected adaptively through the use of learning algorithms, resulting in data that is inherently non-i.i.d. and often lacks desirable statistical properties—such as unbiasedness of simple estimators. A prominent example arises in the stochastic multi-armed bandit setting, where the sample average reward of each arm, when computed naively, is known to exhibit negative bias. Despite this, a precise characterization of such bias under different bandit algorithms has remained elusive.

In this work, we show that data collected by any minimax-optimal bandit algorithm necessarily exhibits non-negligible sample bias. Specifically, we prove that the worst-case bias of the Z-statistic constructed from naive sample means is non-vanishing in the large sample limit. This provides rigorous theoretical justification for the use of debiasing procedures in adaptive data analysis. Furthermore, focusing on the class of UCB-type algorithms, we derive the exact asymptotic leading term of the sample bias and uncover a fundamental trade-off between regret minimization and statistical bias. In particular, we find that despite guarantees of asymptotic normality and unbiasedness in the literature, the bias under standard UCB algorithms (e.g. UCB1) vanishes at an extremely slow rate and remains substantial in practical, finite-time regimes.

4 - Policy Evaluation in Distributional Reinforcement Learning

Lan Wang, University of Miami, Coral Gables, FL, United States

Motivated by the recent success of distributional RL in many practical applications, we study the distributional off-policy evaluation problem in the batch setting when the reward is multi-variate. We propose an offline Wasserstein-based approach to simultaneously estimate the joint distribution of a multivariate discounted cumulative reward given any initial state-action pair in the setting of an infinite-horizon Markov decision process. Finite sample error bound for the proposed estimator with respect to a modified Wasserstein metric is established in terms of both the number of trajectories and the number of decision points on each trajectory in the batch data. I will also discuss non-asymptotic error bounds for online distributional RL. (Joint work with Zhengling Qi, Chenjia Bai, and Zhaoran Wang)

TE56

Building B Level 3 B315

Information in Auctions

Invited Session

Auctions and Market Design

Chair: Yingkai Li, National University of Singapore, Singapore, N/A, Singapore

Co-Chair: Jinzhao Wu, Yale University, 123 York Street, New Haven, CT, 06511, United States

1 - The Role of Prescreening in Auctions with Predictions

Yanwei Sun, Imperial College Business School, London, United Kingdom, Fupeng Sun, Chiwei Yan, Jiahua Wu

Sellers often use closed auctions to create scarcity and prestige, aiming to intensify competition among a select group of high-status bidders. Advances in machine learning and AI make this strategy increasingly viable, enabling cost-effective identification of capable participants. In this paper, we develop a theoretical model to assess whether such practice can be justified from an economic perspective. We consider a setting in which bidders have i.i.d. private valuations, and the auction seller observes a noisy predictor of each bidder's valuation, which is assumed to be fully informative with some probability. Based on this noisy predictor, the seller determines how many bidders to admit -- a process we refer to as *prescreening*. We show that an auction with prescreening is equivalent to a standard auction (i.e., without prescreening) in which bidder valuations are *correlated*. Notably, the standard notion of *affiliation* commonly assumed in the auction literature does not generally hold in this equivalent formulation. We characterize conditions for the existence of symmetric and strictly monotone equilibrium strategies across three classic auction formats: all-pay, first-price, and second-price auctions. Our results demonstrate that prescreening with noisy predictors can significantly enhance revenue in all-pay auctions; in fact, with a perfect predictor, admitting only two bidders is optimal. By contrast, in both first-price and second-price auctions, admitting all bidders remains revenue-maximizing.

2 - Algorithmic Cheap Talk

Konstantin Zabarnyi, Yale University, New Haven, CT, United States, Yakov Babichenko, Inbal Talgam-Cohen, Haifeng Xu

The literature on strategic communication originated with the influential cheap talk model, which precedes the Bayesian persuasion model by three decades. This model describes an interaction between two agents: sender and receiver. The sender knows some state of the world which the receiver does not know, and tries to influence the receiver's action by communicating a cheap talk message to the receiver.

This paper initiates the systematic algorithmic study of cheap talk in a finite environment (i.e., a finite number of states and receiver's possible actions). We first prove that approximating the sender-optimal or the welfare-maximizing cheap talk equilibrium up to a certain additive constant or multiplicative factor is NP-complete. We further prove that deciding whether there exists an equilibrium in which the receiver gets utility higher than the trivial utility he can guarantee is NP-complete. Fortunately, we identify two naturally-restricted cases that admit efficient algorithms for finding a sender-optimal equilibrium - a constant number of states or a receiver having only two actions.

3 - Information Disclosure Makes Simple Mechanisms Competitive

Jinzhao Wu, Yale University, New Haven, CT, United States

In classical mechanism design, the prevailing assumption is that the information structure about agents' types is *exogenous*. This assumption introduces complexity, especially with multi-dimensional agent types, leading to mechanisms that, while optimal, may appear complex and unnatural. Furthermore, Hart and Nisan (2019) show that the gap between the performance of any simple mechanism and the optimal solution could be potentially unbounded. We challenge this conventional view by showing that *simple mechanisms can be highly competitive* if the information structure is *endogenous* and can be influenced by the designer.

We study a multi-dimensional generalization of a single-dimensional model proposed by Bergemann and Pesendorfer (2007), where the designer can shape the information structure via *information disclosure*. Specifically, we consider a fundamental multi-dimensional mechanism design problem, where a seller is selling m items to a single *unit-demand* buyer to maximize her revenue. The buyer's values can be arbitrarily correlated across the items.

Our main result shows that, following an appropriately chosen information disclosure scheme, *item pricing*, i.e., set a take-it-or-leave-it price on each item is highly competitive and guarantees to attain at least 50.1% of the optimal revenue. To our knowledge, this is the first result demonstrating the (approximate) optimality of simple mechanisms in this extensively studied multi-dimensional setting, without making any assumptions about the buyer's value distribution. We believe our result not only demonstrates the power of information disclosure in enhancing the performance of simple mechanisms but also suggests a new framework for reevaluating their efficacy in multi-dimensional settings.

4 - The Simplicity of Optimal Dynamic Mechanisms

Andres Cristi, EPFL, Lausanne, Switzerland, Jose Correa, Laura Vargas Koch

A fundamental economic question is that of designing revenue-maximizing mechanisms in dynamic environments. This paper considers a simple yet compelling market model to tackle this question, where forward-looking buyers arrive at the market over discrete time periods, and a monopolistic seller is endowed with a limited supply of a single good. In the case of i.i.d. and regular valuations for the buyers, Board and Skrzypacz (2016) characterized the optimal mechanism and proved the optimality of posted prices in the continuous-time limit. Our main result considers the limit case of a continuum of buyers, establishing that for arbitrary independent buyers' valuations, posted prices and capacity rationing can implement the optimal anonymous mechanism. Our result departs from the literature in three ways: It does not make any regularity assumptions, it considers the case of general, not necessarily i.i.d., arrivals, and finally, not only posted prices but also capacity rationing takes part in the optimal mechanism. Additionally, if supply is unlimited, we show that the rationing effect vanishes, and the optimal mechanism can be implemented using only posted prices.

TE57

Building B Level 3 B316

Real-World Experimentation: Bridging Academic Insights and Industry Practice

Invited Session

Revenue Management

Chair: Ramesh Johari, Stanford University, Stanford, CA, United States

Co-Chair: Chao Qin, Stanford University, Palo Alto, CA, United States

1 - Shortlisting Using Adaptive Experimentation

Kuang Xu, Stanford Graduate School of Business, Stanford, CA, United States, Chao Qin

Shortlisting involves narrowing a large pool of candidates to a small, high-quality subset likely to include the best option. This challenge appears in domains like experimental design, where cost limits restrict testing to a few treatments from a vast design space. We explore adaptive experimentation to optimize shortlisting and introduce A1 Sampling, a streamlined variant of Thompson Sampling. A1 Sampling offers both robust theoretical guarantees and strong performance in finite-sample simulation experiments.

2 - Best-Arm Identification in a 2 Times 2 Landscape: Frequentist Vs Bayesian Under Fixed-Confidence & Fixed-Budget

Junpei Komiya, NYU Stern, New York, NY, United States

Best-arm identification is the problem of finding which of several arms (options) has the highest expected reward when each arm's reward distribution is unknown and can only be learned through noisy samples. Depending on whether you face a strict sampling budget or aim for a target confidence level and whether you take a frequentist or Bayesian viewpoint, the optimal algorithmic strategy changes. In this talk, we will show how algorithm design is tailored to each of these scenarios.

3 - Pure Exploration via Frank-Wolfe Self-Play

Wei You, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Chao Qin

We study pure exploration in structured stochastic multi-armed bandits, aiming to efficiently identify the correct hypothesis from a finite set of alternatives. For a broad class of tasks, asymptotic analyses reduce to a maximin optimization that admits a two-player zero-sum game interpretation between an experimenter and a skeptic: the experimenter allocates measurements to rule out alternatives while the skeptic proposes alternatives. We reformulate the game by allowing the skeptic to adopt a mixed strategy, yielding a concave-convex saddle-point problem. This viewpoint leads to Frank-Wolfe Self-Play (FWSP): a projection-free, regularization-free, tuning-free method whose one-hot updates on both sides match the bandit sampling paradigm. However, structural constraints introduce sharp pathologies that complicate algorithm design and analysis: our linear-bandit case study exhibits nonunique optima, optimal designs with zero mass on the best arm, bilinear objectives, and nonsmoothness at the boundary. We address these challenges via a continuous-time limit: a differential inclusion with a Lyapunov function that decays exponentially, implying a vanishing duality gap and convergence to the optimal value. Although Lyapunov analysis requires differentiability of the objective, which is not guaranteed on the boundary, we show that along continuous trajectories the algorithm steers away from pathological nonsmooth points and achieves uniform global convergence to the optimal game value. We then embed the discrete-time updates into a perturbed flow and show that the discrete game value also converges. Building on FWSP, we further propose a learning algorithm based on posterior sampling.

4 - Seeing Beyond the Experiment Window: Impacts of Long-Term Ranker Changes

Lo-Hua Yuan, Airbnb, San Francisco, CA, United States, Lei Shi, Peng Ding, Navin Sivanandam

Online randomized experiments are critical for evaluating updates to Airbnb's search ranking algorithm, but short experiment windows limit visibility into longitudinal dynamics: seasonality, evolving user/listing characteristics, and feedback loops between ranker outputs and marketplace evolution. Moreover, past experiment results may not hold in future contexts. To better align ranker deployment decisions with long-term goals, we propose a causal framework to estimate *prospective long-term average treatment effects*: the average effect that a long-term ranker change has on future long-term outcomes. Our approach consists of three components:

1. **Structural nested mean models** (SNMMs) to estimate effects of sustained ranker changes while adjusting for time-varying confounders (e.g., booking intent) and dependencies between past and future ranker exposure;
2. **Surrogate index modeling** using *search result sets* as mediators to impute long-term outcomes (e.g., annual bookings);
3. **Causal transfer learning** to generalize effects from historical data to future ranker evaluation, across temporal shifts (via SNMMs and seasonality patterns) and ranker iterations (via the *result set* surrogate)

Together, these three components offer a way to translate results from short ranker experiments to estimates of future business impacts beyond timescales of any single experiment. This in turn allows decision-makers to better trade off immediate and downstream objectives during ranker iteration, while balancing experimentation constraints. We demonstrate the efficacy of our approach through real-world-based simulation studies, and summarize conditions in which it outperforms alternative estimation strategies.

5 - Beyond Multi-Arm Bandits: Tackling Challenges in Adaptive Experiments at Airbnb

Yicheng Li, Airbnb, San Francisco, CA, United States, Yan Wang, Jiayun Dong

The adoption of adaptive experiments in the tech industry has gained more attention recently. However, current adaptive experiment literature focuses more on improving efficiency of decision-making and less on impact estimation and error control of decision-making. Additionally, adaptive methods explored in research literature usually leverage simplified setups to achieve optimal results even though experiments with real-world impact involve non-stationarity, repeated measure, and infrequent updates, which are often overlooked in literature. Given the complex nature of adaptive experimentation setup for real-world scenarios, classic bandit algorithms are often subject to additional tweaks to improve decision-making, provide impact estimation and provide error control for decision-making - key outcomes of experimentation. At Airbnb, we developed adaptive experiment techniques to tackle these challenges from pre-experiment design and backtesting to post-experiment analysis. In this talk, we will showcase the current practices of adaptive experiments at Airbnb to enhance the efficiency of large-scale online experiments, and share insights for future research and development.

TE58

Building B Level 3 B317

Data-Driven Healthcare Quality and Reliability

Contributed Session

Contributed

Chair: Kazi Arman Ahmed, Oklahoma State University, 716 N Husband St, Stillwater, OK, 74075, United States

1 - The Role of Saliva-Based Testing Programs in Reducing COVID-19 ICU Admissions in Disadvantaged Communities**Alireza Kasaie, Loyola University Chicago, Chicago, IL, United States, Samie Tootooni, Sina Ansari**

The COVID-19 pandemic exposed disparities in severe case management, particularly in disadvantaged communities during the Alpha, Delta, and Omicron waves. This study evaluates the impact of SHIELD Illinois test centers, a saliva-based testing initiative, on COVID-19 ICU admissions in greater Chicago, focusing on socioeconomic factors. We conducted a secondary analysis using data from a medical center's Electronic Health Record (EHR) system and the SHIELD Illinois Testing Program, covering the Alpha, Delta, and Omicron waves. Using linear mixed-effects regression models, we examined the relationship between SHIELD center availability, ICU admissions, center density, and area deprivation index scores. Findings revealed that increased SHIELD center availability during the Omicron wave significantly reduced ICU admissions in disadvantaged zip codes. Each additional SHIELD center reduced ICU admission rates by 0.6% (IRR = 0.994, 95% CI: 0.990–0.998), equating to 0.38 fewer admissions per 1,000 population. Lag analyses indicated further reductions over time, with the incidence rate ratio dropping by 0.7% in underserved areas (IRR = 0.993; 95% CI: 0.989–0.997), translating to 0.45 fewer ICU admissions per 1,000 population. These results underscore the importance of strategically deploying SHIELD test centers to mitigate severe outcomes and promote healthcare equity. This approach offers valuable insights for improving responses to future pandemics in underserved communities.

2 - Mapping Multimorbidity in Neurodegenerative Diseases Through Temporal Network Analysis**Mostafa Amini, California State University Long Beach, Long Beach, CA, United States, Ali Bagheri, Dursun Delen**

Millions of people suffer from neurodegenerative diseases (NDDs) with Alzheimer's alone being the sixth leading cause of death in the US. We study the multimorbidities (i.e., co-occurring diseases) network of NDDs to enhance descriptive and predictive models. These co-occurring conditions may arise independently or from a common biological, social, or environmental antecedent. We use the Cerner data warehouse to study over 25 million encounters of more than 800 thousand patients diagnosed with NDDs. We transform the data into multimorbidity networks, where the nodes are the diseases, and the edges show the existence of a patient diagnosed with both disorders. We construct a series of evolving networks based on patient data from periods before, during, and after the diagnosis of NDDs. These networks grow progressively over time, with increasing numbers of nodes and edges. The pre-diagnosis networks help identify early patterns leading to disease onset, while the post-diagnosis networks capture the development and progression of subsequent comorbidities. This study uses evolutionary multimorbidity networks to explore several predictive tasks: forecasting the appearance of new disease co-occurrences over time, predicting which diseases are likely to emerge in patients or patient groups, estimating the future strength of connections between disease pairs, and analyzing changes in comorbidity clusters. These tasks aim to deepen our understanding of how multimorbidity evolves in the context of NDDs.

3 - Predictive Analysis of Acute Care use Among People with Diabetes**Jinha Lee, Temple University, Philadelphia, PA, United States, Daniel Rubin, Hardik Sharma, Zoran Obradovic**

An estimated 30.2 million Americans have diabetes, and this number is projected to rise. Hospitalizations related to diabetes are common and costly, with approximately one-quarter of adult hospitalizations in the U.S. having a listed diabetes code in recent years. Acute care use, including hospital readmissions and post-discharge Emergency Department (ED) visits, among people with diabetes is a significant driver of healthcare costs and a quality marker. Therefore, focusing acute care use preventive interventions for people at the highest risk would be more efficient and potentially more effective in improving healthcare service outcomes and quality. However, prior research has generally focused on hospital readmissions within 30 days of discharge. While this is an important endpoint, it excludes hospitalizations that occur more than 30 days after discharge, all ED visits, and people who have not been hospitalized.

This research focuses on a way to predict the risk of hospitalizations and ED visits over a longer period (1 year) of individuals with diabetes. Using electronic health records of 22,666 patients with encounters, demographic factors, vital records, diagnosis records, prescribing history, and lab test results, this research predicts the acute care use over a 1-year period and identifies its relevant importance features. The result of this study could provide a better understanding of risk factors for and preventive factors against acute care use, and eventually guide the selection of the best interventions for individuals according to their risk profile.

4 - Predicting Long Length of Hospital Stays from Clinical Notes and Electronic Health Records**Shahrazad Khanizadeh, University at Buffalo, Buffalo, NY, NY, United States, Justin Kwok, Haimonti Dutta**

Accurate prediction of Length of Stay (LoS) can enhance hospital resource management, reduce readmissions, and improve patient care. In recent years, there has been a widespread adoption of electronic health records resulting in the availability of clinical notes from patients. In addition, significant advances in machine learning (including deep and reinforcement learning), natural language processing, and related statistical methods have facilitated accurate predictions of LoS. Surprisingly, the distribution of LoS in hospitals tends to have a very high variance -- some patients stay for less than one day while others can be in the hospital for several months. In this research, we take a closer look at what drives Long Length of Stay (LLOS) by examining clinical notes and extracting from them factors that can extend stays in the hospital. In doing so, we closely observe managerial aspects of the stay such as hospital course and acute, chronic and transitional issues related to it, need to be transferred between departments or ICU for treatment, types of service required (surgery, medicine, neurology, orthopedics, etc.), discharge condition and disposition, etc. After extracting these features from the clinical notes, we integrate them with multimodal data extracted from EHRs and build models to effectively predict LLOS.

5 - Comorbidity Patterns in Jia-Diagnosed Patients**Kazi Arman Ahmed, Oklahoma State University, Stillwater, OK, United States, Hamid Hadidi, Mohammad Fili, Dursun Delen, Guiping Hu, Lizhi Wang**

Juvenile idiopathic arthritis (JIA) is a chronic pediatric inflammatory disease often accompanied by other health conditions. However, the timing of these comorbidities and variations across patient subgroups require more in-depth investigation. JIA is relatively uncommon but significant, affecting approximately 1 to 2 in every 1,000 U.S. children and resulting in an estimated 7,000–10,000 new diagnoses each year. In this study, we leverage Oracle Real-World Data, a large-scale electronic health record database from diverse U.S. healthcare settings, to analyze comorbidity patterns among pediatric JIA patients both before and after diagnosis. Key objectives include: 1) identifying the demographic characteristics of children diagnosed with JIA; 2) characterizing common pre-diagnostic comorbid conditions that might serve as early clinical indicators; and 3) profiling post-diagnostic comorbidities to understand how the disease's burden evolves over time. Understanding these temporal comorbidity trends and demographic differences may inform improved early screening protocols and comprehensive long-term care strategies for JIA patients.

TE59

Building B Level 3 B318

Theory and Applications of Analytics

Contributed Session

Contributed

Chair: Jiaqi Qiu, Plantage Muidergracht 12, 1018TV

1 - Mixed-Integer Programming and Constraint Programming Approaches for Precast Concrete Flowshop Scheduling with Multiple Production Lines

Siyan Wang, Huazhong University of Science and Technology, Wuhan, China, People's Republic of, Nanfang Cui

Timely production scheduling and delivery of precast concrete components are critical for prefabricated building projects. This study addresses a precast flowshop scheduling problem that involves customized molds and common production pallets across multiple lines. A mixed-integer programming (MIP) model is developed, which effectively solves small-scale problems to optimality. To tackle the inherent complexity, this paper introduces a constraint programming (CP) model leveraging interval decision variables and constraint propagation algorithms. Experimental results demonstrate that the CP model outperforms both the MIP model and existing algorithms in small-scale instances and real-world case studies. Notably, the MIP and CP models have proved the optimality of solutions for 38 out of 60 small-scale instances. The findings highlight the potential of the CP model as a powerful tool for solving scheduling problems efficiently and underscore the importance of adopting exact algorithms for scheduling in precast production.

2 - Analyzing Makespan-Sojourn Time Trade-Offs for Open Shop Schedule with Unit-Time Operations

Gwanhee Lee, Pohang University of Science and Technology, Pohang, Korea, Republic of, Kangbok Lee, Michael Pinedo

We consider the open shop scheduling problem under a maximum makespan constraint, where each job consists of unit-time operations, each processed on a different machine. The problem is structurally similar to other combinatorial optimization problems, such as timetabling and edge coloring of complete bipartite graphs. To evaluate schedule performance, we focus on the sojourn time of a job, defined as the elapsed time between a job's start time and completion time. Minimizing sojourn time enhances system efficiency and reduces operational costs associated with prolonged job processing. However, there is an inherent trade-off between sojourn time and makespan, since allowing a longer makespan can decrease job sojourn times. In this study, we analyze the trade-off relationship between makespan and sojourn time.

3 - Deep Nonnegative Matrix Factorization with Beta Divergences

Akwum Onwunta, Lehigh University, Bethlehem, PA, United States

Deep Nonnegative Matrix Factorization (deep NMF) has recently emerged as a valuable technique for extracting multiple layers of features across different scales. However, all existing deep NMF models and algorithms have primarily centered their evaluation on the least squares error, which may not be the most appropriate metric for assessing the quality of approximations on diverse datasets. For instance, when dealing with data types such as audio signals and documents, it is widely acknowledged that beta-divergences offer a more suitable alternative. In this talk, we present new models and algorithms for deep NMF using some beta-divergences, with a focus on the Kullback-Leibler divergence. Subsequently, we apply these techniques to the extraction of facial features, the identification of topics within document collections, and the identification of materials within hyperspectral images.

4 - Neglecting Initial Population Uncertainty Leads to Biased Distributional Flood Risk Outcomes

Parin Bhaduri, Cornell University, Ithaca, NY, United States, Adam Pollack, Vivek Srikrishnan

Efforts to improve urban resilience rely on understanding population-level impacts and responses to environmental shocks. These analyses are subject to a number of critical uncertainties, which have the potential to greatly influence coupled natural-human system outcomes. For example, assumptions around population demographics can influence the capacity for adaptation and the distribution of risk and damages. One approach to explore the impact of demographic assumptions is the use of synthetically-generated populations. However, synthetic population generation often relies on sampled survey data, such as from the U.S. Census, which yield highly uncertain estimates of household demographic characteristics. Neglecting such uncertainties could result in biased projections of future environmental impacts, but studies that utilize synthetic populations are typically silent on the sensitivity of model outcomes to these initial population conditions. Here, we quantify the impact of initial population uncertainty along with other key uncertainties, such as model parameters and exogenous forcings, on simulations of agent-based urban-scale flooding with an agent-based model. After calibrating our model to openly available data, we identify how uncertainties influence a variety of model outputs over the course of a multi-decadal simulation.

5 - Uncovering the Drivers of Centenarian Lifespans: a Data Analytical Approach

Samaneh Rezaeimanesh, Oklahoma State University, Stillwater, OK, United States, Mohammad Fili, Guiping Hu, Alex Bishop

As the global population ages, exceptional longevity, defined as living to 100 years or more, is becoming an increasingly important subject in demographic and aging research. Understanding the factors that contribute to such extended lifespans is essential not only for addressing the evolving needs of aging populations, but also for uncovering the mechanisms that enable some individuals to reach such old age. This study employs analytical methods to uncover the pathways to exceptional longevity. This form of longevity can be influenced by a range of factors, including demographic, familial, social, occupational, and lifestyle characteristics. This study investigates the effects of these elements using a dataset of 1858 centenarians. Our findings may offer valuable insights into the life course experiences that could shape extended lifespans, informing strategies to support aging populations and enhance healthy aging.

6 - Explainable AI Signal Diagnosis for AI-Based Statistical Process Monitoring

Jiaqi Qiu, University of Amsterdam, Amsterdam, Netherlands, Rob Goedhart, Jannis Kurtz, Inez Maria Zwetsloot

AI techniques have excelled in many fields, including statistical process monitoring (SPM). Although existing AI-based SPM methods show a promising detection ability in high-dimensional scenarios, post-signal diagnosis for black-box models is often ignored. The lack of explainability makes it difficult to make an out-of-control action plan and take appropriate actions after the signal. Explainable AI (XAI) techniques, like feature attribution and counterfactual explanations, can help decision-makers investigate these signals. Feature attribution helps explain which features contribute to the output most, and counterfactual explanations provide insights into which minimal changes to features (perturbations) make the difference between an in-control or out-of-control observation. This paper proposes a framework integrating XAI signal diagnosis with AI monitoring methods. The applicability of the proposed framework is validated by the experiments generating explanations on simulated and real data.

TE60

Building B Level 3 B319

Machine Learning Theory and Applications

Contributed Session

Contributed

Chair: Chi Chun Chien, National Chi Nan University, Nantou County, N/A, Taiwan

1 - Machine Learning in the Mortgage Industry: a Literature Review

Sajede Farahani, University of Wisconsin-Milwaukee, Milwaukee, WI, United States, Wilkistar Otieno

Credit risk refers to the potential financial loss incurred when a borrower fails to meet their contractual repayment obligations. Effectively managing credit risk is a fundamental component of risk management for financial institutions. In recent decades, credit risk assessment—particularly in consumer lending—has become increasingly data- and model-driven. Financial institutions now rely heavily on credit scoring models to predict the likelihood of delinquency or default across a range of consumer credit products, including mortgages, auto loans, credit cards, and personal loans.

In this study, we present a literature review of existing research methods and machine learning (ML) techniques applied specifically to mortgage credit scoring models, with a focus on how Explainable Artificial Intelligence (XAI) can enhance the robustness and transparency of predictive modeling. We find that only a limited number of studies have incorporated XAI methods into mortgage credit scoring, highlighting a significant gap in the current literature. Moreover, although many individual studies include macroeconomic indicators—such as interest rates, unemployment rates, and inflation—as well as climate-related variables, there is a notable absence of any comprehensive framework that organizes and classifies these variables. To address this shortcoming, we propose a structured categorization intended to serve as a practical reference for researchers and practitioners seeking to incorporate these factors more systematically into future credit risk models.

2 - Interpretable Electrode-Band Optimization for Affective Emotion Decoding

Jorge Fresneda Fernandez, New Jersey Institute of Technology, Newark, NJ, United States, Sumya Akter, Simon Bin Akter, Md Mohit Hasan, Tanmoy Sarkar Pias, David Eisenberg, Chelsea Hill, Jerry Fjermestad

Emotion decoding is pivotal for advancing neurotechnology and affective computing. Prior electrode selection methods often relied on statistical or spatial heuristics, limiting interpretability and cross-dataset scalability. This study incorporates an explainable AI framework employing SHapley Additive exPlanations (SHAP) to identify the most informative electrodes and frequency bands for emotion decoding. Signals were first converted from time to frequency domain using Fourier Transform. A residual CNN trained on the DEAP dataset with 32 brain electrodes and 8 additional physiological signals outperformed 16 popular models across Valence, Arousal, Dominance, and Liking (precision: 0.97, specificity: 0.96, sensitivity: 0.98, F1-score: 0.97), exceeding baselines by 32.54%. SHAP-based feature selection (threshold ≥ 0.02) identified an optimal 25-electrodes (61 features), retaining 98.78% of full-model (200 features) performance while reducing input parameters by 23.93%. Cross-dataset generalization on SEED-V and AMIGO showed strong consistency, with 21-mapped electrodes (52 features) and 11-mapped electrodes (32 features) reaching 96.25% and 98.04% similarity to the full 64-electrode (320 features) and Emotive 17-electrode (85 features) setups for detecting six (Happiness, Sadness, Fear, Anger, Disgust, Neutral) and four (Valence, Arousal, Dominance, Liking) emotions, while reducing parameters by 35.86% and 9.79%, respectively. Key insights reveal a dominant role of fronto-central and parieto-occipital brain regions in Delta and High Beta bands. Additionally, non-brain signals—especially from eye activity, facial muscle engagement, and respiration—emerged as influential under Delta band analysis, underscoring their often-overlooked value in emotional decoding. This interpretable, optimized, and scalable approach reduces computational complexity, offering an efficient solution for real-time emotion recognition in neurotechnology applications.

3 - Machine Unlearning via Information Theoretic Regularization

Shizhou Xu, University of California Davis, Davis, CA, United States, Thomas Strohmer

How can we effectively remove or “unlearn” undesirable information, such as specific features or individual data points, from a learning outcome while minimizing utility loss and ensuring rigorous guarantees? We introduce a mathematical framework based on information-theoretic regularization to address both feature and data point unlearning. For feature unlearning, we derive a unified solution that simultaneously optimizes diverse learning objectives, including entropy, conditional entropy, KL-divergence, and the energy of conditional probability. For data point unlearning, we first propose a novel definition that serves as a practical condition for unlearning via retraining, is easy to verify, and aligns with the principles of differential privacy from an inference perspective. Then, we provide provable guarantees for our framework on data point unlearning. By combining flexibility in learning objectives with simplicity in regularization design, our approach is highly adaptable and practical for a wide range of machine learning and AI applications. We demonstrate its effectiveness across diverse data modalities, including tabular data, images, and large language models, through comprehensive numerical experiments.

4 - Modular and Adaptive Conformal Prediction for Two-Stage Models via Residual Decomposition

William Zhang, MIT Operations Research Center, Cambridge, MA, United States, Saurabh Amin, Georgia Perakis

Conformal prediction offers finite-sample coverage guarantees under minimal assumptions. However, existing methods treat the entire modeling process as a blackbox, overlooking opportunities to exploit modular structure. While model-aware conformal techniques have been proposed for specialized architectures, conformal prediction for modular two-stage blackbox pipelines remains underexplored. We introduce a conformal prediction framework for two-stage models, where an upstream predictor generates intermediate representations for a downstream model. Such pipelines are common in domains like macroeconomic forecasting and supply chain modeling, where interpretability and retrainability are crucial for downstream decision-makers. By decomposing the overall residual into stage-specific components, our method enables interpretable uncertainty attribution and modular risk control, facilitating more informed and accountable decisions. To ensure valid coverage, we calibrate stage-wise scaling parameters using a risk-controlling selection procedure based on family-wise error rate (FWER), and extend this to an adaptive variant for online settings. Our method provides long-run coverage guarantees and improves robustness to distribution shifts. Experiments on synthetic upstream, downstream, and pipeline-wide shifts, as well as real-world economic indicator forecasting, show that our approach maintains coverage over Adaptive Conformal Inference (ACI) and proportional-integral-derivative control (PID) and provide stage-aware, interpretable prediction intervals support that robust decision-making.

5 - Nonparallel Support Vector Ordinal Regression for Imbalanced Data

Chi-Chun Chien, Department of Information Management, National Chi Nan University, Nantou, Taiwan, Puli, Taiwan, Jing-Rung Yu, Stephan Jao, Donald Lien, Pi-Hsia Hung, Hou-shen Chen

Support Vector Ordinal Regression (SVOR) is a multiclass classification method designed for ordered categories by constructing parallel discriminant hyperplanes. However, conventional SVOR struggles with imbalanced datasets, often exhibiting bias toward majority classes and resulting in poor classification accuracy for minority classes. To address this limitation, we propose an enhanced approach that integrates accuracy sensitivity into a nonparallel SVOR framework, referred to as NSVORA. This method employs an ℓ_1 -norm support vector machine with accuracy-sensitive mechanism to construct class-specific hyperplanes independently, allowing each class to have its own decision boundary. Additionally, the model incorporates constraints that maximize the lower bounds of the true positive rate (TPR) and true negative rate (TNR) for each class, thereby improving classification performance across all classes. By removing the restriction of parallel hyperplanes, NSVORA effectively addresses the shortcomings of conventional SVOR in handling class imbalance. The model was evaluated with datasets from the UCI Machine Learning Repository as well as other sources, including data on Taiwanese companies to predict financial distress. Empirical results demonstrate that the proposed model significantly enhances classification performance, particularly for minority classes.

TE61

Building B Level 4 B401

Computational and Statistical Methods for Uncertainty Quantification

Invited Session

Quality, Statistics and Reliability

Chair: Moses Chan, Northwestern University, Evanston, IL, 60208, United States

Co-Chair: Naichen Shi, University of Michigan, Ann Arbor, MI, United States

1 - Surrogate Active Subspaces for Jump-Discontinuous Functions

Nathan Wycoff, University of Massachusetts Amherst, Amherst, MA, United States

Surrogate modeling and active subspaces have emerged as powerful paradigms in computational science and engineering. Porting such techniques to computational models in the social sciences brings into sharp relief their limitations in dealing with discontinuous simulators, such as Agent-Based Models, which have discrete outputs. Nevertheless, prior applied work has shown that surrogate estimates of active subspaces for such estimators can yield interesting results. But given that active subspaces are defined by way of gradients, it is not clear what quantity is being estimated when this methodology is applied to a discontinuous simulator. We begin this article by showing some pathologies that can arise when conducting such an analysis. This motivates an extension of active subspaces to discontinuous functions, clarifying what is actually being estimated in such analyses. We also conduct numerical experiments on synthetic test functions to compare Gaussian process estimates of active subspaces on continuous and discontinuous functions. Finally, we deploy our methodology on Flee, an agent-based model of refugee movement, yielding novel insights into which parameters of the simulation are most important across 8 displacement crises in Africa and the Middle East.

2 - Efficient Gaussian Process Modeling for Replicated and Censored Data in Manufacturing Applications

Emily Kang, University of Cincinnati, Cincinnati, OH, United States, Ying Zhang, Shuo Li

Gaussian process (GP) regression is widely used for modeling responses in both physical and computer experiments. In practice, data from physical experiments may include replications and be subject to censoring due to equipment limitations or contextual constraints. Motivated by real-world manufacturing data, we develop a GP modeling framework that efficiently analyzes replicated and potentially censored data. Our approach leverages recent advances in likelihood-based inference and latent-

variable methods for GPs, effectively exploiting replication while rigorously incorporating censoring information—analogue to exact simulation techniques for truncated distributions. We demonstrate the effectiveness of our method on synthetic manufacturing data, showing that it provides enhanced prediction, uncertainty quantification, and computational scalability, making it well-suited for large-scale applications with structured replication and incomplete data.

3 - Bayesian Bridge Gaussian Process Regression

Lulu Kang, University of Massachusetts Amherst, Amherst, MA, United States

Gaussian Process (GP) regression is a widely used surrogate modeling approach for both computer and physical experiments, especially when modeling complex and highly nonlinear input-output relationships. However, like many supervised learning methods, GP regression can suffer from the "curse of dimensionality." The goal of this project is to perform shrinkage regularization to both the regression coefficients of the mean part of the GP and the length-scale of the correlation function. Under the Bayesian framework, it is similar to the Bayesian bridge regression, which leads to constrained posterior distribution of the parameters. We propose using the Spherical Hamiltonian Monte Carlo method to solve the constrained sampling problem. Through several examples, we have shown the proposed method outperforms other alternative in prediction accuracy and variable selection.

4 - Probabilistic Reduced Order Models

Naichen Shi, Northwestern University, Evanston, IL, United States, Yuheng Wang, Wenxi Cai

As computer simulation models grow increasingly sophisticated, they also become more computationally expensive. Reduced-order models (ROMs) offer a computational feasible solution by simplifying high-dimensional simulation problems and operating in a reduced latent space, with results later mapped back to the physical domain. Traditional ROMs rely on dimension reduction techniques such as Singular Value Decomposition (SVD) and variational autoencoders (VAEs) to establish this latent-physical correspondence. However, these methods often suffer from limited expressiveness and struggle to reproduce complex stochastic dynamics in the original physical space.

In this work, we propose a novel framework that leverages diffusion models to more effectively bridge the latent and physical spaces. Diffusion models offer a powerful generative capability, allowing us to capture rich, multi-modal distributions and reconstruct physically realistic dynamics. By integrating simple physical mechanisms with the learning of denoising neural networks, our approach produces predictions that are both statistically plausible and consistent with governing physical principles. We demonstrate the effectiveness of our method using two case studies: fluid flow simulation and laser-based additive manufacturing.

TE62

Building B Level 4 B402

Technometrics Invited Session

Invited Session

Quality, Statistics and Reliability

Chair: Kamran Paynabar, ISyE Georgia Tech, Atlanta, GA, United States

1 - An Adaptive Sampling Strategy for Real-Time Anomaly Detection with Unmanned Sensing Vehicles

Yue Jiang, Purdue University, West Lafayette, IN, United States, Ana Maria Estrada Gomez

Unmanned sensing vehicles (USVs) have been widely used for real-time anomaly detection in various applications, including environmental monitoring, precision agriculture, and military surveillance. The USVs collecting data can only provide partial information about the space being monitored. Thus, it is critical to decide where to deploy the USVs at each point in time to maximize the change detection capability, while minimizing deployment costs. This article proposes an adaptive sampling strategy for real-time anomaly detection with USVs. First, a novel spatio-temporal sequential tensor decomposition algorithm is developed to decompose the high-dimensional data collected by the USVs into three components, a spatial component, a temporal component, and a sparse component, that captures the locations suspicious of change. The spatial and temporal components are used for one-step prediction to guide the adaptive sampling strategy. The strategy is designed to maximize the detection power and control the deployment costs. The main idea is to balance exploration and exploitation by designing a sampling distribution function to decide where to collect data at each acquisition time. The movement of the USVs is controlled by using Voronoi tessellations on the sampling distribution function. The performance of the proposed framework is demonstrated through simulations and case studies.

2 - Personalized Tucker Decomposition: Modeling Commonality and Peculiarity on Tensor Data

Hao Yan, Arizona State University, Chandler, AZ, United States, Kamran Paynabar, Jiuyun Hu, Naichen Shi, Raed Al Kontar

We propose a personalized Tucker decomposition (perTucker) to address the limitations of traditional tensor decomposition methods in capturing heterogeneity across different datasets. perTucker decomposes tensor data into shared global components and personalized local components. We introduce an order orthogonality assumption and develop a proximal gradient regularized block coordinate descent algorithm guaranteed to converge to a stationary point. The unique and common representations learned by perTucker reveal intrinsic statistical patterns in data and provide valuable information for a wide range of downstream analytics, including anomaly detection, source classification, and clustering. We demonstrate perTucker's effectiveness through a simulation study and two case studies on solar flare detection and tonnage signal classification.

3 - Convolutional Non-Homogeneous Poisson Process and Its Application to Wildfire Ignition Risk Quantification for Power Delivery Networks

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

This article proposes a spatio-temporal point process model, known as the Convolutional Non-homogeneous Poisson Process (cNHPP), and applies the model to quantify wildfire ignition risks for power delivery networks. The proposed model captures both the current (i.e., instantaneous) and cumulative (i.e., historical) effects of key environmental processes (i.e., covariates) on wildfire risks, as well as the spatio-temporal dependency among different segments of the power delivery network. The computation and interpretation of the intensity function are thoroughly investigated. We apply the proposed approach to estimate wildfire ignition risks on major transmission lines in California, using historical fire data, meteorological and vegetation data obtained from the National Oceanic and Atmospheric Administration and National Aeronautics and Space Administration.

4 - Targeted Variance Reduction: Effective Bayesian Optimization of Black-Box Simulators with Noise Parameters

Simon Mak, Duke University, Durham, NC, United States

The optimization of a black-box simulator over control parameters arises in broad scientific applications. Here, the simulator often takes the form $f(x, \theta)$, where θ are parameters that are uncertain in practice. Stochastic optimization aims to optimize the objective $E[f(x, \theta)]$, where $\theta \sim P$ is a random variable that models uncertainty on θ . For this, existing black-box methods typically employ a two-stage approach for selecting the next point (x, θ) , where x and θ are optimized separately via different acquisition functions. As such, these approaches do not employ a joint acquisition over (x, θ) , and thus may fail to fully exploit control-to-noise interactions for effective stochastic optimization. We propose a new Bayesian optimization method called Targeted Variance Reduction (TVR). The TVR leverages a novel joint acquisition function over (x, θ) , which targets variance reduction on the objective within the desired region of improvement. Under a Gaussian process surrogate on f , the TVR acquisition can be evaluated in closed form, and reveals an insightful exploration-exploitation-precision trade-off for stochastic black-box optimization. The TVR can further accommodate a broad class of non-Gaussian distributions on P via a careful integration of normalizing flows. We demonstrate the improved performance of TVR over the state-of-the-art in a suite of numerical experiments and an application to the robust design of automobile brake discs under operational uncertainty.

TE63

Building B Level 4 B403

Recent Advances in Learning, Prediction, and Data-Driven Decision Making

Invited Session

Data Mining

Chair: Yu Ma, University of Wisconsin Madison, Madison, WI, United States

1 - Sparse Plus Low-Rank Decomposition of Large Language Models

Mehdi Makni, MIT, Cambridge, MA, United States, Xiang Meng, Rahul Mazumder

Sparse plus Low-Rank (S+LR) decomposition of Large Language Models (LLMs) is emerging as a promising approach in model compression (reducing the size of neural networks), where one aims to decompose pre-trained model weights into a sum of sparse and low-rank matrices ($W \sim S + LR$). This decomposition can be an appealing alternative to pure sparsification of models (setting parameters to zero) under the same compression ratio. It is also Hardware-aware as it can result in inference acceleration on modern GPUs (compared to the dense model). Despite recent progress, existing methods suffer from substantial performance degradation relative to dense models. In this work, we introduce 3BASiL, a highly efficient one-shot post-training method for (S + LR) decomposition of LLMs that addresses this gap. Our approach first leverages a novel 3-Block Alternating Direction Method of Multipliers (ADMM) approach, termed 3BASiL, to minimize the layer-wise reconstruction error with convergence guarantees. Our numerical experiments show that 3BASiL reduces the WikiText2 perplexity gap to dense model by over 35% compared to prior methods for a Llama3.2-3B model under a 50%-Sparse plus 128-low-rank configuration. Moreover, our method achieves faster compression runtime than current state-of-the-art (S + LR) decomposition techniques. The compression of a Llama3.2-3B model under a 50%-Sparse plus 128-low-rank configuration takes ~1.5 hours on a single L40 48GB GPU.

2 - Learning to Price Rentals Under Unknown Degradation and Demand

Joo Seung Lee, UC Berkeley, Berkeley, CA, United States, Anil Aswani

We study a dynamic pricing problem in equipment rental where equipment degrades with use and incurs a cost upon failure. The renter receives only binary feedback after each use as to whether the equipment fails or survives, along with contextual information about rental usage-intensity and cumulative past usage. The goal is to estimate the latent degradation dynamics from interaction data and use this knowledge to inform pricing decisions. This work also lays the foundation for learning demand-side behavior, specifically the distribution of customer willingness-to-pay, toward designing an optimal rental pricing policy that balances revenue and risk over the equipment's lifetime.

3 - Flow-Based Conformal Prediction for Multi-Dimensional Time Series

Junghwan Lee, Georgia Institute of Technology, Atlanta, GA, United States, Chen Xu, Yao Xie

Conformal prediction for time series presents two key challenges: (1) leveraging sequential correlations in features and non-conformity scores and (2) handling multi-dimensional outcomes. To address these challenges, we propose a novel conformal prediction method for time series using Flow with classifier-free guidance. Specifically, we use a flow to model the distribution of prediction residuals and their transformations conditioned on historical context, which is encoded using Transformer. Non-conformity scores are defined as the Euclidean distance between transformed prediction residuals and the mean of a Gaussian source distribution, enabling the construction of prediction sets directly in the source distribution space at a target confidence level. We provide theoretical guarantees that our method satisfies marginal coverage and derive finite-sample bounds for conditional coverage under stationary features. Experiment results on real-world multi-dimensional time series datasets demonstrate that our method constructs smaller prediction sets while satisfying a target coverage, outperforming existing baselines.

4 - Data to Dose: Personalized Dosing with Data Generation and Expert Feedback

Hansraj Satyam Verma, Carnegie Mellon University, Pittsburgh, PA, United States, Holly Wiberg, Woody Zhu, Sridhar Tayur

Personalized treatment optimization is often constrained by data scarcity, particularly in settings involving heterogeneous or rare patient profiles. We introduce **GenEx**, a hybrid framework that integrates preference-based Bayesian optimization with a generative model that is fine-tuned online using expert feedback. Expert-provided pairwise rankings are used to update both the surrogate model and the generative component, which proposes synthetic labels in uncertain regions when agreement with expert preferences is high. The expert is selectively queried in regions of high model uncertainty. Our model accommodates imperfect expert feedback and variation in source data quality. Theoretical analysis establishes sublinear regret bounds that account for bias introduced by synthetic augmentation. In a numerical study simulating a development environment, GenEx achieves faster convergence and improved decision quality compared to preference-only and purely generative baselines with a smaller starting data set. This demonstrates the promise of combining generative modeling with expert feedback for principled decision-making in low-data regimes.

TE64

Building B Level 4 B404

Humanitarian Operations

Contributed Session

Contributed

Chair: Naimur Rahman Chowdhury, North Carolina State University, 2151 Avent Ferry Rd Apt 207, Raleigh, NC, 27606, United States

1 - Integrating Heterogeneous Suffering In Disaster Response: a Robust Estimation and Optimization Framework

Yuxin Hong, Tianjin University, Tianjin, China, People's Republic of, Ning Zhu

The increasing frequency and severity of natural disasters present significant challenges for effective humanitarian logistics, particularly in the aftermath of the disaster. In this paper, we propose a joint vehicle routing and supply allocation model that integrates an extended area-level Deprivation Cost Function (DCF) to mitigate human suffering. The DCF is estimated using survey data, and we incorporate regression residuals to capture the heterogeneous suffering among affected populations, thereby formulating the problem as a stochastic program. To address estimation uncertainties inherent in survey data and the potential ambiguity in residual distribution, we propose a robust satisficing model that accounts for these uncertainties and approximates it as a finite-dimensional program. Exploiting the finite-valued and maximization-based structure of our cost function, we devise an L-shaped method with a tailored algorithm to solve the subproblem, and we further introduce a conservative approximation to enhance computational scalability. Extensive numerical experiments validate the algorithm's efficiency and demonstrate the model's robustness in mitigating estimation errors and distribution ambiguities. The comparison of different strategies reveals that accounting for heterogeneous suffering leads to more balanced and effective relief operations. Additionally, strategically prioritizing aid to less resilient areas, even at the cost of moderate distribution imbalances, significantly reduces extreme suffering without compromising overall fairness.

2 - Operationalizing Spatial Food Insecurity Using Cell Phone Mobility Data: a Network-Based Alternative to Usda's Food Access Research Atlas

Dwight Lewis Jr, The University of Alabama, Tuscaloosa, AL, United States, Irem Sengul Orgut, Lauren Davis, Shonna Morgan

Food insecurity—limited or uncertain access to nutritious food—affects millions across the U.S., with accessibility as a key determinant. Spatial access to grocery stores offering fresh, healthy ingredients supports health-protective behaviors. Traditional tools like the USDA's Food Access Research Atlas (FARA) rely on uniform, distance-based measures between supply and demand points, which often fail to reflect actual consumer behavior. Without a full picture of communities' food environments, this limitation can undermine the ability of public health and hunger relief organizations to develop targeted and effective food access interventions.

This research introduces a network-based alternative to FARA using 2022–2024 cell phone mobility data from Dewey Data's SafeGraph dataset. Our methodology employs two key innovations: (1) network analysis of consumer movement patterns to delineate empirically derived grocery service areas, and (2) spatial accessibility measurement within those areas.

We analyze mobility data across all 50 states and Washington, D.C., focusing on supermarkets (NAICS 445110), meat/farmers markets (445210, 445230), and warehouse clubs/supercenters (445231). Using network analysis grounded in information theory, we delineate grocery service areas and apply floating catchment methods to examine accessibility within these zones—enabling the identification of potential grocery deserts in context of the local food environment.

As a next step, we will simulate optimal grocery market placements by targeting high-traffic convenience stores (NAICS 445120) and dollar stores (452319) as potential sites to qualify for a hypothetical fresh food financing policy. While ongoing work remains, this framework offers stakeholders a behavior-based understanding of food access that complements and enhances FARA's traditional, distance-based approach.

3 - A People-Centric Approach to Model Risks in Post-Wildfire Recovery and Resource Allocation Using A Two-Stage Machine Learning Modeling

Poulomee Roy, State University of New York at Buffalo, Buffalo, NY, United States, Sayanti Mukherjee

In recent years, the frequency and severity of wildfires have increased across the United States due to anthropogenic interferences, poor land management, and Wildland-Urban Interface (WUI) expansion. Although predicting wildfires has improved, challenges remain in wildfire suppression and resource allocation meeting people's needs. This project aims to develop a people-centric approach that accounts for socio-economic characteristics, demographics, living conditions of WUI residents, and wildfire impact to identify risks to post-wildfire recovery and guide resource allocation for wildfire mitigation.

We categorized the high- and low-risk cohorts based on the wildfire impacts in the first stage of the proposed two-stage machine learning (ML) modeling framework. We implemented ML classifiers (e.g., random forest (RF), boosting) and selected CatBoost as the most effective classifier, achieving a predictive accuracy of 95%. In the second stage, we investigated each risk cohort separately to understand how the risk of scarce resource allocation is associated with the wildfire severity, socioeconomic characteristics, and the demographic composition of the affected population, leveraging multivariate linear regression, ridge and lasso regression, RF, decision tree, XGBoost, and Gradient Boosting, achieving 90% accuracy across the cohorts.

Our research shows that lower-income individuals, those residing in congested housing, the disabled, the educationally disadvantaged, and the elderly individuals are most affected, lacking adequate resources to cope with wildfire impacts. The insights from our analysis will help policymakers design an efficient people-centric rescue plan that addresses supply and targeted resource allocation while considering people's needs and reducing delays in rescue operations.

4 - Bi-Level Optimization Integrating Socioeconomic and Geospatial Analysis for Disaster Response Facility Allocation

Vinayaka Gude, Elon University, Elon, NC, United States

This study presents a multi-level analytical framework for optimizing the placement of disaster recovery centers integrating socio-economic vulnerability, physical accessibility, and storm damage severity. The initial model employs a population-based greedy algorithm to select facility locations that maximize risk-weighted coverage and minimize access disparity. In the second phase, a road network analysis is performed to assess the spatial accessibility of each facility. The model accounts for connectivity within the drivable network and elevation impacts, identifying the top three locations with the highest reachability and least terrain impediments. A damage-weighted efficiency metric is introduced to capture how well each facility serves the most affected counties. Results from Western North Carolina after Hurricane Helene show that the model-selected locations consistently outperform the real deployed disaster recovery locations across all metrics, offering broader and more equitable service to high-risk populations. By combining socio-demographic, infrastructural, and geographic considerations, this approach supports more just and operationally feasible disaster response planning.

5 - Building Capacity Function with Dynamic Categorization of Food Agencies in Hunger-Relief Networks

Naimur Rahman Chowdhury, North Carolina State University, Raleigh, NC, United States, Julie Ivy, Osman Ozaltin

Estimating the capacity for receiving donations (CRD) in a geographic region is a crucial step in effective and equitable donation allocation in food bank networks. The CRD in a geographic region is influenced by the operational performance of food agencies within the region, which changes over time. Therefore, estimating CRD requires a dynamic approach instead of exploring the empirical distribution over all temporal periods. In this study, we develop a dynamic approach to estimate CRD by categorizing agencies into different groups with dynamic features. To choose the best lookback window for dynamic agency feature extraction, we use multiple recurrent neural networks (RNNs). Then, we use partitional and hierarchical unsupervised learning for agency categorization. Furthermore, we use a non-parametric approach to obtain the probability density function (PDF) of donations received in each agency category. Finally, we develop an aggregated sampling method to develop a capacity function for a region with multiple agencies. We illustrate results from Food Bank of Central and Eastern North Carolina (FBCENC) data and show significant statistical differences in the capacity function for regions with different compositions of agencies.

TE65

Building B Level 4 B405

Recent Advances in Data-Driven Decision-Making and Analytics

Invited Session

Artificial Intelligence

Chair: Yueying Li, Johns Hopkins Carey Business School, Baltimore, MD, United States

Co-Chair: Ali Eshragh, Johns Hopkins Carey Business School, Washington, DC, United States

1 - Winner's Curse in Data-Driven Decision-Making: Evidence and Solutions

Sikun Xu, Washington University in St. Louis, St. Louis, MO, United States, Raphael Thomadsen, Dennis Zhang

Data-driven decision-making involves estimating the value associated with each possible decision and selecting the optimal estimated choice. This type of decision making is at the heart of a huge number of modern marketing applications, including ad creative choice, algorithm optimization, personalized targeting, A/B testing, pricing, and assortment optimization. In practice, it is crucial not only to estimate the optimal policy but also to accurately measure the incremental value or lift of that policy. In this paper, we first demonstrate theoretically that selecting the optimal policy based on estimated effects from data leads, on average, to overly optimistic evaluations of the policy value, a phenomenon known as the winner's curse. This is true no matter what best methodology is used to estimate the policy's effectiveness. We then empirically illustrate that the winner's curse arises in a wide range of key marketing applications, including A/B testing, personalized targeting, and counterfactual estimation using structural models, and that its magnitude can be substantial even within realistic parameter ranges commonly seen in the literature. Given the generality of this problem across diverse settings, we propose a correction method based on a non-continuous bootstrap approach designed to effectively mitigate the winner's curse in nearly all scenarios. Finally, we benchmark our proposed method against several existing context-specific solutions, demonstrating that our bootstrap-based correction consistently performs well and frequently outperforms alternative methods across important marketing contexts.

2 - Learning to Cooperate Without Using Non-Public Information

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

We consider price competition among multiple sellers over a selling horizon of T periods. In each period, sellers simultaneously offer their prices and subsequently observe their respective demand that is unobservable to competitors. The realized demand of each seller depends on the prices of all sellers following a private unknown linear model. We propose a manipulatable pricing algorithm, which does not require sellers to communicate demand information or coordinate prices once the selling horizon starts. With property algorithm inputs, our algorithm leads to competitive prices, while with manipulated algorithm inputs, our algorithm leads to cooperative prices. We show that our algorithm, when employed by all sellers, leads at a fast convergence rate $O(1/\sqrt{T})$ to the competitive/cooperative prices that sellers would reach if they were fully informed. Particularly, in the manipulated case, sellers achieve an optimal order-of- $\sqrt{T}$ regret relative to the static benchmark policy of cooperative prices as well as a linear revenue increment relative to the static benchmark policy of competitive prices.

3 - Data-Driven Knowledge Transfer in Batch Q^* Learning

Wenbo Jing, New York University, New York, NY, United States, Elynn Chen, Xi Chen

In data-driven decision-making across marketing, healthcare, and education, leveraging large datasets from existing ventures is crucial for navigating high-dimensional feature spaces and addressing data scarcity in new ventures. We investigate knowledge transfer in dynamic decision-making by focusing on batch stationary environments and formally defining task discrepancies through the framework of Markov decision processes (MDPs). We propose the Transfer Fitted Q -Iteration algorithm with general function approximation, which enables direct estimation of the optimal action-state function Q^* using both target and source data. Under sieve approximation, we establish the relationship between statistical performance and the MDP task discrepancy, highlighting the influence of source and target sample sizes and task discrepancy on the effectiveness of knowledge transfer. Our theoretical and empirical results demonstrate that the final learning error of the function is significantly reduced compared to the single-task learning rate.

4 - Robustness Versus Statistical Efficiency: Superiority of Naive Optimization via Worst-Case Stochastic Dominance

Arindam Roy Chowdhury, Columbia University, New York City, NY, United States, Henry Lam

Many decision-making tasks give rise to optimization problems with uncertainty in the underlying parameter or probability distribution. This uncertainty can arise from statistical noises due to limited data, or non-stationary parameter shifts that are not predictable from the past. In the literature, various approaches ranging from (distributionally) robust optimization to regularization have been proposed, as attempts to obtain solutions that perform well statistically or robustly against model shifts. Contrary to common belief, we show that, in terms of the behavior of regret or optimality gap, a naive optimization formulation based on empirical or simple point estimate of the uncertain parameter is in a sense optimal, regardless of whether the uncertainty comes from statistical noises, unpredictable shifts, or both. Our superiority assertion for naive optimization is based on a new notion of worst-case stochastic dominance for the asymptotic regret, and is argued via what we call the "symmetry" of uncertainty that applies universally over statistical errors, adversarial errors, and their interactions. We show how our results apply to the comparisons among a wide range of existing data-driven and robust optimization formulations.

5 - Scalable Data-Driven TIME Series Analytics with Applications in the Energy Sector

Ali Eshragh, Johns Hopkins Carey Business School, Washington, DC, United States, Yuexing Li, Luke Yerbury, Fred Roosta, Michael Mahoney

We develop a new scalable algorithm for approximating leverage scores in large datasets using techniques from randomized numerical linear algebra. Leverage scores play a central role in tasks such as data sampling, outlier detection, and dimensionality reduction, making their efficient estimation critical for modern analytics applications.

Building on this, we introduce a fast and interpretable algorithm for modeling large-scale time series data. Together, these methods enable data-driven modeling and decision-making in resource-intensive settings.

We demonstrate the effectiveness of our approach through extensive experiments, including applications in the energy market, where large volumes of high-frequency data demand scalable and accurate forecasting tools. Our methods achieve up to 800x speedups over exact approaches while maintaining high accuracy, offering practical solutions for time series analytics in energy systems and beyond.

TE66

Building B Level 4 B406

Clustering, Motifs, and Spatial Discovery in Complex Data

Contributed Session

Contributed

Chair: Li Zhang, University of Texas Rio Grande Valley, Edinburg, TX, United States

1 - Clustering Us Metro Areas by Per-Capita Mobile Greenhouse Gas Emissions: Targeted Mitigation Strategies

Peiyao Zhao, University of Massachusetts, Amherst, MA, United States, Jimi Oke

We present a novel typology of 349 metropolitan statistical areas (MSAs) in the United States and demonstrate its implication for understanding and mitigating per-capita mobile greenhouse gas emissions (PMGE). Via factor analysis of 55 indicators across 7 categories of road network, buildings, socioeconomic, car ownership, commute mode share, transit energy, and climate, we identified eight key drivers of mobile emissions: Auto + Scale, Transit Propensity, Density, Sustainable Transit, Warm Weather, Remote + Active, Network Connectivity, Network Sprawl. Based on these, we clustered the metro areas via Gaussian mixture modeling to obtain seven metro types, namely Auto Large, Auto Small, Dense Active, Heavyweight Eco-Transit, Heavyweight Transit, Sparse Active, and Sparse Auto. We used an Extreme Gradient Boosting regression model to predict PMGE based on all the indicators, excluding population. By using SHapley Additive exPlanations (SHAP) values, we identified the most significant indicators for all MSAs and for each cluster type.

The model reveals that density and transit mitigate PMGE, while car use and climate exacerbate it. The typology provides further insights into how these effects vary across network, mode share and development patterns. Ultimately, the typology can serve as a framework to inform and initialize type-specific mitigation strategies.

2 - How to Get the Most Accurate Results for Your Data Using Ensembles for Numeric, Classification, Logistic and TIME Series Data

Russ Conte, ensembles1.com, Forest Park, IL, United States

Ensembles are well established as a way to get the most accurate results for data. This presentation will highlight four new packages that automatically produce a wide range of individual models and ensembles of models for numeric data, classification data, logistic data, and time series data.

If you want to see how to get the best possible results for your data, this session is for you!

The session will begin with an introduction to building ensembles, so everyone starts from the same point. It is NOT necessary to have prior ensemble building experience, that's what the packages will do for you automatically.

A very important point will be demonstrating how the packages address the issue of reproducibility of results. Each of the four packages have built-in methods to produce results on untrained data that are testable, falsifiable, and fully reproducible with very high accuracy.

The packages contain many other features which can be used to provide the most accurate results.

This is a technical session, and will cover the steps used to get optimal solutions, such as how the NumericEnsembles package can be used to eliminate correlated predictors, how the ClassificationEnsembles package automatically produces 20 confusion matrices, how the LogisticEnsembles package can be used for sports and HR analytics, and how the ForecastingEnsembles package can be used to predict labor market data.

Many examples will be demonstrated in real time to allow you to see how you can use the power of ensembles in your work.

3 - Estimating Motif Counts via Generalized Intersection Local Times

Qinghua Ding, UC Berkeley, Berkeley, CA, United States, Venkat Anantharam

We study the problem of estimating structural properties of graphs using local random walk observations. Quantities such as triangle density and motif counts are fundamental in network analysis, but computing them exactly is often infeasible for large graphs due to memory and computational constraints. Prior work has explored random-walk-based estimators for specific statistics, but these are often limited to simple cases like vertex or triangle counts and may lack theoretical guarantees on sample complexity.

A general framework is proposed based on generalized intersection local times, which measure how often multiple independent random walks simultaneously visit the vertices of a target subgraph pattern. This formulation allows us to construct estimators for a broad class of subgraph counts in a unified and principled way.

We analyze the variance and sample complexity of these estimators under standard assumptions. For many random graph models, our approach achieves sublinear sample complexity in the graph size. Importantly, our estimators require only local neighborhood access, making them practical for massive graphs such as social networks.

This work establishes a scalable, general-purpose, and theoretically grounded framework for subgraph count estimation. We also outline ongoing extensions to other local graph statistics and to sample-efficient machine learning based on data gathered via random walks on large networks.

This work is supported in part by NSF grants CCF-1901004 and CIF-2007965.

4 - Comparison of Clustering Analysis Methods on Behavioral Supply Chain Data

Souri Sasanfar, Northeastern University, Boston,, MA, United States, Jacqueline Griffin

This study explores whether different clustering methodologies yield distinct insights into behavioral patterns observed in pharmaceutical supply chain simulations. We applied both single-phase (Longest Common Prefix (LCP) and Dynamic Time Warping (DTW)) and two-phase clustering approaches (LCP-LCP and DTW-DTW, segmented into pre- and post-disruption periods) to gameplay data from 169 participants. Analysis of cluster compositions revealed notable discrepancies: behaviors classified distinctly by LCP—such as 'Followers,' 'Reactors,' and 'Hoarders'—mapped differently onto DTW clusters, indicating complex behavioral distinctions between methods. Statistical validation using Chi-Square tests, Cramér's V, and Cohen's Kappa confirmed significant associations and differences across clustering methods. Our findings support hypotheses that clustering methods substantially influence the interpretation of behavioral data. Specifically, disruptions and information availability significantly alter ordering behaviors, underscoring the importance of methodological choice in behavioral analysis. These results highlight how leveraging multiple clustering techniques can provide comprehensive behavioral insights critical for improving decision-making frameworks in complex, disrupted environments

5 - A Theory-Informed Text Mining Approach to Identifying and Classifying Strategic Paradoxes in Organizational and Sectoral Contexts

Gulce Cuhaci Tasdelen, Istanbul Bilgi University, STAMFORD, CT, United States

Conceptualized as “contradictory yet interrelated elements that exist simultaneously and persist over time” (Smith & Lewis, 2011), paradoxes are especially salient in decision-making contexts where leaders must navigate competing demands and dual transformation pressures (Smith, Binns & Tushman, 2010). Strategic paradoxes, persistent and interdependent tensions such as exploration vs. exploitation or control vs. flexibility, are increasingly recognized as central to contemporary organizational complexity. However, identifying and classifying such tensions within large-scale organizational and sectoral texts remains a methodological challenge. This paper introduces a theory-informed text mining approach that detects and classifies paradoxical expressions in strategic documents through rule-based regular expressions (regex), dynamic paradox typologies, and natural language processing (NLP) techniques. The framework is grounded in paradox theory, drawing on foundational models by Smith and Lewis (2011) and extended by research on paradoxical leadership (Lavine, 2014; Zhang et al., 2015), mindset and cognition (Miron-Spektor et al., 2017; Boemelburg et al., 2023), and institutional and processual perspectives (Jarzabkowski et al., 2013; Besharov & Smith, 2014; Schad et al., 2016). We operationalize paradox detection through a curated inventory of “seed expressions” and regex patterns, applied to the U.S. healthcare and insurance sector reports, and aim to integrate sectoral and transformational tagging via rule-based and semantic similarity models. Findings demonstrate the framework’s potential for sectoral benchmarking, longitudinal text analysis, and future integration into leadership analytics. We also discuss its applicability for studying paradoxical leadership behaviors and paradoxical mindset development, offering scalable pathways to support both academic research and practical decision-making in complex environments.

6 - Towards Robust Time Series Modeling via Motif Discovery

Li Zhang, University of Texas Rio Grande Valley, Edinburg, TX, United States, Salomon Ibarra

Deep learning models have achieved remarkable progress in recent years and have the potential to make a transformative impact on critical applications such as renewable energy and disaster response. However, current deep learning methods mostly rely on average loss -- which often excels in average performance and fails to deliver reliable predictions in critical tasks — the samples occurs less yet important and difficult to predict. Such discrepancy in performance greatly hurts the reliability and usability of deep learning models in the real world. To address this problem, we propose we propose a new framework incorporates time series motif behavior into training framework. We will demonstrate the effectiveness and efficiency through our experiment results.

TE67

Building B Level 4 B407

Recent Advances in Tensor Decomposition and Completion Algorithms and Their Applications

Invited Session

Data Mining

Chair: Mehrdad Ghadiri, MIT, Cambridge, United States

Co-Chair: Xinyu Chen, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Approximately Optimal Core Shapes for Tensor Decompositions

Matthew Fahrbach, Google Research, Atlanta, GA, United States, Mehrdad Ghadiri, Gang Fu, Vahab Mirrokni

This work studies the combinatorial optimization problem of finding an optimal core tensor shape, also called multilinear rank, for a size-constrained Tucker decomposition. We give an algorithm with provable approximation guarantees for its reconstruction error via connections to higher-order singular values. Specifically, we introduce a novel Tucker packing problem, which we prove is NP-hard, and give a polynomial-time approximation scheme based on a reduction to the 2-dimensional knapsack problem with a matroid constraint. We also generalize our techniques to tree tensor network decompositions. We implement our algorithm using an integer programming solver, and show that its solution quality is competitive with (and sometimes better than) the greedy algorithm that uses the true Tucker decomposition loss at each step, while also running up to 1000x faster.

2 - Quantifying Time Series Periodicity with Interpretable Machine Learning

Xinyu Chen, Massachusetts Institute of Technology, Cambridge, MA, United States

Time series autoregression is a classical statistical model for capturing auto-correlations and identifying temporal patterns such as periodicity and seasonality. In this work, we propose a novel sparse autoregression framework from an interpretable machine learning perspective and the model interpretability for periodicity quantification is reinforced by L0-norm induced sparsity constraints. On the multidimensional time series that involves complicated spatial and temporal dimensions, we propose a spatially- and time-varying sparse autoregression model and resolve the mixed-integer programming problem by developing a two-stage optimization scheme. In particular, the proposed scheme makes the model scalable to large problems even with millions of decision variables. Empirically, we conduct extensive experiments to evaluate the proposed model on real-world time series data. First, we apply our framework to real-world human mobility data in NYC and Chicago, and our analysis of ridesharing data from 2019 to 2024 demonstrates the disruptive impact of the COVID-19 pandemic on mobility regularity and the subsequent recovery trends. Second, we demonstrate the spatial patterns of yearly seasonality in climate variable time series across the past four decades, and our model allows to discover dynamic climate patterns and identify climate phenomena such as El Nino in sea surface temperature. Thus, our framework lays an insightful foundation for understanding periodicity in real-world time series.

3 - Matrix Completion with Cross-Concentrated Sampling: Bridging Uniform Sampling and Cur Sampling

HanQin Cai, University of Central Florida, Orlando, FL, United States

While uniform sampling has been widely studied in the matrix completion literature, CUR sampling approximates a low-rank matrix via row and column samples. Unfortunately, both sampling models lack flexibility for various circumstances in real-world applications. In this work, we propose a novel and easy-to-implement sampling strategy, coined Cross-Concentrated Sampling (CCS). By bridging uniform sampling and CUR sampling, CCS provides extra flexibility that can potentially save sampling costs in applications. In addition, we also provide a sufficient condition for CCS-based matrix completion. Moreover, we propose a highly efficient non-convex algorithm, termed Iterative CUR Completion (ICURC), for the proposed CCS model. Numerical experiments verify the empirical advantages of CCS and ICURC against uniform sampling and its baseline algorithms, on both synthetic and real-world datasets.

4 - Understanding the Challenges of Entry-Wise Matrix Completion Under Arbitrary Observation Patterns

Ruijia Cao, Cornell University, Ithaca, NY, United States

Matrices are fundamental units for representing and organizing data: estimating missing entries from the often noisy and incomplete observed ones therefore comprises a crucial task in modern data science. While the field of matrix completion has been well-studied, most existing techniques rely on strong distributional or structural assumptions on the observation pattern. In many real world applications, however, the observation patterns may be highly non-uniform, thus limiting the applicability of existing approaches.

In this work, we consider the special case where the ground-truth matrix is a rank-1 sign matrix and the observed entries are random flips of the true ones. Given a specific target entry to estimate, we show that the min-max optimal estimator of any missing entry is given by the Bayes estimator under a uniform prior, which can also be expressed as a function of an appropriate Eulerian flow on the observation graph. We then illustrate how non-uniformity in the observations could result in the failure of existing matrix completion algorithms such as the spectral method. These insights lead to improved algorithms that explicitly account for the non-uniformity in observations.

TE68

Building B Level 4 B408

AI and RL for Robust and Adaptive Decision-Making in Business

Invited Session

Data Mining

Chair: Elynn Chen, New York University, New York, NY, United States

1 - Transfer Reinforcement Learning: Value-Based Methods for Non-Stationary Mdp

Elynn Chen, New York University, New York, NY, United States

In dynamic decision-making scenarios across business, healthcare, and education, leveraging data from diverse populations can significantly enhance reinforcement learning (RL) performance for specific target populations, especially when target samples are limited. We develop comprehensive frameworks for transfer learning in RL, addressing both stationary Markov decision processes (MDPs) with iterative Q*-learning and non-stationary finite-horizon MDPs with backward inductive Q*-learning. For stationary MDPs, we propose an iterative Q*-learning algorithm with knowledge transfer, establishing theoretical justifications through faster convergence rates under similarity assumptions. For non-stationary finite-horizon MDPs, we introduce two key innovations: (1) a novel "re-weighted targeting procedure" that enables cross-satege transfer along multiple temporal steps, and (2) transfer deep Q*-learning that leverages neural networks as function approximators. We demonstrate that while naive sample pooling strategies may succeed in regression settings, they fail in MDPs, necessitating our more sophisticated approach. We establish theoretical guarantees for both settings, revealing the relationship between statistical performance and MDP task discrepancy. Our analysis illuminates how source and target sample sizes impact transfer effectiveness. The framework accommodates both transferable and non-transferable transition density ratios while assuming reward function transferability. Our analytical techniques have broader implications, extending to supervised transfer learning with neural networks and domain shift scenarios. Empirical evidence from both synthetic and real datasets validates our theoretical results, demonstrating significant improvements over single-task learning rates and highlighting the practical value of strategically constructed transferable RL samples in both stationary and non-stationary contexts.

2 - Doubly high-dimensional contextual bandits: An interpretable model for joint assortment-pricing

Linda Zhao, University of Pennsylvania, Philadelphia, PA, United States

3 - Online Estimation and Inference for Robust Policy Evaluation in Reinforcement Learning

Yichen Zhang, Purdue University, West Lafayette, IN, United States

Reinforcement learning has emerged as one of the prominent topics attracting attention in modern statistical learning, with policy evaluation being a key component. Unlike the traditional machine learning literature on this topic, our work emphasizes statistical inference for the model parameters and value functions of reinforcement learning algorithms. While most existing analyses assume random rewards to follow standard distributions, we embrace the concept of robust statistics in reinforcement learning by simultaneously addressing issues of outlier contamination and heavy-tailed rewards within a unified framework. In this paper, we develop a fully online robust policy evaluation procedure, and establish the Bahadur-type representation of our estimator. Furthermore, we develop an online procedure to efficiently conduct statistical inference based on the asymptotic distribution. This paper connects robust statistics and statistical inference in reinforcement learning, offering a more versatile and

reliable approach to online policy evaluation. Finally, we validate the efficacy of our algorithm through numerical experiments conducted in simulations and real-world reinforcement learning experiments.

4 - A Customer-Aware Recommender System for Influencer Marketing

Xuan Bi, University of Minnesota, Minneapolis, MN, United States

Influencer marketing is transforming digital retail by leveraging social media influencers to promote products. However, matching the right influencers with right products remains challenging. This is because influencers act as intermediaries, and a proper matching should reflect not only influencers' own tastes, but also the underlying preferences of their following customers. To address this, we take a design science approach and propose a Customer-Aware Recommender system for Influencer marketing (CARI), an AI-powered recommender framework that integrates customer behavior into influencer-product matching. CARI combines graph neural networks and self-attention mechanisms, enhanced with customer-aware embeddings, to capture the multifaceted preferences and interactions among influencers, products, and customers. An online field experiment on a leading influencer platform in Asia showed that CARI significantly improves upstream influencer endorsement engagement by 105% and downstream customer spending by 19.8 times.

TE69

Building B Level 4 B409

Preference Learning in Large Language Models

Invited Session

Data Mining

Chair: Chong Liu, State University of New York at Albany, Albany, NY, United States

Co-Chair: Fan Yao, University of North Carolina at Chapel Hill, Chapel Hill, United States

1 - Statistical Impossibility and Possibility of Aligning LLMs with Human Preferences: from Condorcet Paradox to Nash Equilibrium

Weijie Su, Wharton School, University of Pennsylvania, Philadelphia, PA, United States

Aligning large language models (LLMs) with diverse human preferences is critical for ensuring fairness and informed outcomes when deploying these models for decision-making. We seek to uncover fundamental statistical limits concerning aligning LLMs with human preferences, with a focus on the probabilistic representation of human preferences and the preservation of diverse preferences in aligned LLMs. We first show that human preferences can be represented by a reward model if and only if the preference among LLM-generated responses is free of any Condorcet-cycle. Moreover, we prove that Condorcet cycles exist with probability converging to one exponentially fast under a probabilistic preference model, thereby demonstrating the impossibility of fully aligning human preferences using reward-based approaches such as reinforcement learning from human feedback. Next, we explore the conditions under which LLMs would employ mixed strategies--meaning they do not collapse to a single response--when aligned in the limit using a non-reward-based approach, such as Nash learning from human feedback (NLHF). We identify a necessary and sufficient condition for mixed strategies: the absence of response that is preferred over all others by a majority. We prove that this condition holds with high probability under the probabilistic preference model, thereby highlighting the statistical possibility of preserving minority preferences without explicit regularization in aligning LLMs. Finally, we leverage insights from our statistical results to design a novel, computationally efficient algorithm for finding Nash equilibria in aligning LLMs with NLHF. Our experiments show that Llama-3.2-1B, aligned with our algorithm, achieves a win rate of 60.55% against the base model.

2 - Competition and Diversity in Generative AI

Manish Raghavan, Massachusetts Institute of Technology, Cambridge, MA, United States

Recent evidence suggests that the use of generative artificial intelligence reduces the diversity of content produced. In this work, we develop a game-theoretic model to explore the downstream consequences of content homogeneity when producers use generative AI to compete with one another. At equilibrium, players indeed produce content that is less diverse than optimal. However, stronger competition mitigates homogeneity and induces more diverse production. Perhaps more surprisingly, we show that a generative AI model that performs well in isolation (i.e., according to a benchmark) may fail to do so when faced with competition, and vice versa. We validate our results empirically by using language models to play Scattergories, a word game in which players are rewarded for producing answers that are both correct and unique. We discuss how the interplay between competition and homogeneity has implications for the development, evaluation, and use of generative AI.

3 - Understanding the Preferences of Large Language Models Under Interdisciplinary Contexts

Yushun Dong, Florida State University, Tampa, FL, United States

This presentation explores the preferences and behaviors of large language models (LLMs) when applied to interdisciplinary tasks. We first introduce Political-LLM, a structured framework for integrating LLMs into political science research, bridging positive and normative traditions via taxonomy-guided alignment. Through a voting simulation case study, we uncover ideological preferences in LLM outputs, emphasizing the need for domain adaptation and ethical safeguards. Complementing this, we present CEB, a compositional evaluation benchmark that systematically assesses bias in LLMs across multiple social dimensions and tasks. By combining the two studies, we reveal how LLMs encode and express context-dependent preferences, highlighting risks such as ideological skew and social stereotyping. This interdisciplinary investigation provides new insights into value alignment and responsible deployment of LLMs.

4 - Faster Wind: Accelerating Iterative Best-of-N Distillation for LLM Alignment

Bo Dai, Georgia Tech, Atlanta, GA, United States

Recent advances in aligning large language models with human preferences have corroborated the growing importance of best-of-N distillation (BOND). However, the iterative BOND algorithm is prohibitively expensive in practice due to the sample and computation inefficiency. In this talk, we address the problem by revealing a unified game-theoretic connection between iterative BOND and self-play alignment, which unifies seemingly disparate algorithmic paradigms. Based on the connection, we establish a novel framework, WIN rate Dominance (WIND), with a series of efficient algorithms for regularized win rate dominance optimization that approximates iterative BOND in the parameter space. We provide provable sample efficiency guarantee for one of the WIND variants with the square loss objective. The experimental results confirm that our algorithm not only accelerates the computation, but also achieves superior sample efficiency compared to existing methods.

5 - Coactive Learning for Large Language Models Using Implicit User Feedback

Kianté Brantley, Harvard University, Allston, MA, United States

We propose coactive learning as a model and feedback mechanism for training large language models (LLMs). The key insight is that users provide implicit feedback whenever they edit the text y proposed by an LLM. While the edited text $\bar{y}$ is typically not a gold-standard example for supervised training, coactive learning merely requires that the edited text $\bar{y}$ is an improvement over the proposed text y . Note that such weak implicit preference feedback $\bar{y} \succcurlyeq y$ is available in many application settings on a per-user basis, thus enabling the personalization of LLMs. In this paper, we develop the theoretical basis for coactive training of non-linear models, and we derive CoRLL as the first coactive learning algorithm for LLMs. Empirical results indicate that CoRLL is effective even for weak and noisy coactive preference feedback, making it a promising algorithm for training and personalization of LLMs from feedback that is naturally collected in many use cases.

Tuesday, October 28, 5:45 PM - 6:45 PM

Building A Level 4 A411

Omega-Rho Lecture

Plenary Keynote

Keynote

Chair: Tao Hong

1 - AI and OR in Action: Enhancing Decision-Making for Forest Fire Risk Mitigation

Andres Weintraub, Universidad de Chile, Santiago, Chile

Climate change has made forest fires more frequent and destructive. Increasingly preventive measures, taken before fires start, are becoming essential. In particular, fuel management, like fire breaks, where areas that allow fire to spread are cleared out to stop or delay the advance of fires. This Omega Rho Distinguished Lecture will describe AI and OR tools we have developed to support fire prevention decisions. The extremely high uncertainty in fire ignitions and spread require complex approaches. Particularly challenging problems are how to detect and prevent intentional fires and how simulation and optimization models support decisions on where firebreaks are more effective in stopping the growth of megafires.

Building B Level 2 B206

Keynote - The Irrational Decision How We Gave Computers the Power to Choose for Us

Plenary Keynote

Keynote

1 - The Irrational Decision How We Gave Computers The Power to Choose for US

Benjamin Recht, University of California Berkeley, Berkeley, CA, United States

This talk traces the intellectual history of automated decision-making to its origins in the post-World War II development of computers. Mathematicians of the 1940s set out to design machines that could act as ideal rational agents in the face of uncertainty. In this pursuit, a cluster of foundational mathematical technologies—including linear programming, game theory, and neural networks—emerged as a foundation for a mathematical rationality, which changed how we think about human decision-making itself.

As the operations research community knows, these seminal computational methods have grown into a robust discipline and industry with countless success stories. I'll describe several such successes in accelerating computers, regulating pharmaceuticals, and deploying electronic commerce. These examples will highlight how automated decision systems excel in specific sweet spots with clear rules, well-defined goals, and well-constrained contexts.

However, most of life, for better or for worse, doesn't fit into neat mathematical models. Even in our data-driven world, we make decisions based on values, emotions, and incomplete information that can't be reduced to optimization problems. I will thus close by exploring how to best harness eighty years of unfathomable computational progress while preserving human agency and judgment.

Building A Level 4 A412

Edelman Reprise

Plenary Keynote

Keynote

Chair: Irv Lustig, Princeton Consultants, Coral Gables, FL, 33146, United States

1 - Edelman Reprise: Usa Cycling

Ryan Cooper, USA Cycling, Georgetown, TX, United States, Jim Miller

In Olympic cycling, the difference between gold and missing the podium is measured in fractions of a second. For USA Cycling's Women's Team Pursuit squad, the 2023 World Championships in Glasgow highlighted a stark reality: a 6th-place finish with a time of 4:12.684 and missing the qualification round for the bronze medal final by 0.159 seconds. With the Paris 2024 Olympics approaching and the team falling to an 8th-place World Ranking, the challenge was clear: reduce nearly seven seconds off their performance in just one year, an improvement considered nearly impossible in elite sports. With limited funding compared to powerhouse nations like Great Britain and Germany, USA Cycling relied on Operations Research (OR), Machine Learning (ML), data analytics, race simulation, and targeted athlete development to bridge the gap. Armed with cutting-edge analytics and modeling, real-time performance tracking, and aerodynamic innovations, this initiative aimed at optimizing every aspect of the team's race strategy. Through the power of data-driven decision-making, USA Cycling achieved what few thought was possible: a stunning eight-second reduction in time to capture Olympic gold, with a time of 4:04.32.

Wednesday, October 29, 8:00 AM - 9:15 AM

WA01

Building A Level 3 A301

From Data to Decision: Intelligent Decision-Making with Applications in Manufacturing and Healthcare

Invited Session

Quality, Statistics and Reliability

Chair: Jihoon Chung, Hanyang University, Seoul, N/A, Korea, Republic of

Co-Chair: Ziyang Zhang, Oklahoma State University, Stillwater, OK, 74078, United States

1 - Development of a Rubber Property Prediction Model Based on Integrated use of Mixing Process Sensor and Recipe Data

Jung-Hwan Hong, Pusan National University, Geumjeong-Gu, Korea, Republic of, Sa-Eun Park, Ho-Jin Cho, Chie-Hyeon Lim, Hye-Mee Kim, Hyeong-Woo Choi, Jik-Hyun Yoon, Ki-Hun Kim

The rubber mixing process is a critical stage in rubber manufacturing, where multiple raw materials are combined according to a predefined recipe to achieve desired material properties. However, product quality is assessed through physical and chemical tests conducted only after the mixing is complete, and the results are typically available several hours later. Consequently, any quality issues in the compounded rubber are identified only after the fact, limiting timely intervention. To address this challenge, we propose a predictive model that estimates key quality characteristics of rubber at the completion of the mixing process. The model integrates heterogeneous data sources — real-time process sensor data and recipe data — both of which are essential for accurate prediction. Experiments conducted on real-world industrial data demonstrate that the proposed model significantly outperforms state-of-the-art benchmarks in terms of prediction accuracy. By enabling real-time estimation of rubber properties, the model offers potential for early defect detection and enhanced process reliability in rubber manufacturing.

2 - TestImputer: Neural Imputation to Mitigate Information Loss in Reduced Semiconductor Wafer Probe Test

Min-Kun Shin, Pusan National University, Pusan, Korea, Republic of, Sa-Eun Park, Young-Mok Bae, Seung-Hyun Choi, Ki-Hun Kim

Semiconductor Wafer Probe Test (WPT) verifies each chip's electrical integrity, but running thousands of subtests per chip greatly increases test costs. Test Reduction (TR) based on chip sampling alleviates this burden but introduces Unsampled Chip Measurements (UCMs) that undermine multivariate quality analytics and elevate test escape risk. This study proposes TestImputer, a graph neural imputation network that imputes UCMs accurately while mitigating information loss due to TR. TestImputer first transforms tabular WPT data into a graph structure by representing chips and subtests as nodes and assigning their measurements as attributes of the edges connecting chip and subtest nodes. Simultaneously, TestImputer connects each chip to its adjacent chips to utilize Adjacent Chip Correlation (ACC), where spatially adjacent chips exhibit similar electrical measurements due to locally shared process conditions and gradual variations across the wafer surface. A attention mechanism then weights each neighboring chip, propagating information to impute the UCMs of the target chip. Experiments on real-world WPT data under different TR scenarios, characterized by unsampled chip ratios of 25%, 50%, and 75%, demonstrate that TestImputer consistently outperforms state-of-the-art baselines, achieving average improvements of 3.7% in root mean squared error, 3.0% in mean absolute error, and 4.8% in Wasserstein distance. TestImputer enables manufacturers to reduce WPT costs while mitigating the information loss.

3 - A Machine Learning-Based Reduced Order Model for Predicting Aerosol Dynamics in Dry Powder Inhalers

Ziyang Zhang, Oklahoma State University, Stillwater, OK, United States, Chenang Liu, Yu Feng

Accurately modeling aerosol transport in dry powder inhalers (DPIs) is critical for understanding and improving drug delivery efficiency. Traditional high-fidelity Computational Fluid Dynamics-Discrete Element Method (CFD-DEM) simulations provide detailed insights into particle-flow interactions and emitted particle size distributions. However, CFD-DEM simulations are computationally intensive and impractical for real-time or patient-specific optimization.

This work presents a data-driven Reduced Order Model (ROM) that leverages machine learning (ML) to fast predict emitted aerodynamic particle size distributions (APSDs), dosage metrics and patient usage conditions. A CFD-DEM simulation framework is used to generate training data, incorporating parametric sweeps across inhaler geometries, drug formulations, and patient-specific inhalation profiles.

The ROM architecture integrates autoencoders with regression networks and attention mechanisms to capture both global aerosol dynamics and localized particle interactions. Data augmentation strategies, including parametric perturbation and symmetry-based transformations, are employed to enrich training diversity and improve model generalizability.

Preliminary validation shows that the ROM achieves agreement with CFD-DEM simulations on key aerosol metrics, including emitted APSDs and spatial dose distributions. By significantly reducing simulation time while maintaining predictive accuracy, the proposed ROM provides a promising framework for real-time DPI design evaluation and personalized respiratory drug delivery modeling.

4 - Diffusion Model-Enhanced Integrative Approach for Adaptive Tumor Segmentation in Ultrasound Imaging

Emmanuel Yangué, Oklahoma State University, Stillwater, OK, United States, Chenang Liu, Yuxuan Li, Ashish Ranjan

Accurate identification of tumor regions, referred to as regions of interest (ROIs), in ultrasound imaging is critical for advancing image-guided drug delivery (IGDD). However, segmentation remains challenging due to high levels of interference and noise. To address these challenges, this study introduces IC-DREAM, an adaptive framework that integrates intensity-based centroid estimation for improved ROI center identification and employs a diffusion-based model to enhance boundary delineation, thereby improving segmentation accuracy and quality. The algorithm was validated on both simulated and real-world ultrasound image datasets, showing superior accuracy and reliability over conventional methods, all without the need for labeled data or training. While IC-DREAM effectively segments ROIs from ultrasound images, its performance remains highly dependent on the accurate identification of ROI centers and contours. These findings underscore the potential of IC-DREAM to tackle complex challenges in ultrasound-based tumor segmentation.

WA02

Building A Level 3 A302

Optimization and Scheduling in Complex Logistics Systems

Contributed Session

Contributed

Chair: Tzvika Geft, Rutgers University, New Brunswick, NJ, United States

1 - Improving Picking Efficiency in E-Commerce Warehouses: a Correlation-Aware Scattered and Commingled Storage Optimization

TAO HE, University of Science and Technology of China, Hefei, China, People's Republic of

With the rapid expansion of e-commerce, warehouse picking efficiency has become a decisive competitive factor. Responding to the “small-batch, high-frequency, and SKU-diverse” nature of e-commerce orders, this study focuses on the **Scattered & Commingled Storage Assignment (SCSA)** strategy and, for the first time, systematically incorporates SKU demand correlations into this framework. We define groups of SKUs that frequently co-occur in the same order as **Demand Correlation Patterns (DCPs)** and formulate a mixed-integer programming model that maximizes the spatial proximity of correlated SKUs while preserving SKU dispersion constraints. To tackle the resulting NP-hard problem, we develop a two-stage heuristic algorithm based on fix-and-optimize framework that achieves tight alignment with a theoretical lower bound, thereby ensuring computational efficiency. Computational experiments and simulation studies show that our correlation-aware approach outperforms traditional scattered storage schemes, and that commingled storage configurations offer additional operational advantages. The analysis yields actionable insights into dynamic replenishment policies and slot-capacity settings, striking an optimal balance between inventory availability and accessibility. By resolving the inherent conflict between SKU dispersion and correlation clustering, this research advances warehouse optimization for modern e-commerce operations and provides a practical solution for picker-to-part systems handling complex order profiles.

2 - Ilp-Based Preemption Decisions in Twin Yard Crane Scheduling with Bidirectional Flows

Kim Jeong Heon, Pusan National University, Busan, Korea, Republic of

As global maritime container volumes continue to rise, efficient container handling at ports is becoming increasingly vital. Yard cranes—particularly in Twin Yard Crane Systems (TYCS), where two cranes operate on a shared rail—are crucial for improving terminal throughput. However, the non-crossing constraint between cranes adds significant scheduling complexity. This study focuses on the preemption decision problem in TYCS, where the goal is to determine whether and where to preempt container transport tasks to reduce interference and enhance efficiency. Unlike previous studies that assume unidirectional flows or fixes preemption points, we address a more realistic setting of bidirectional container flows. We propose a novel approach that approximates a single-crane schedule using an Eulerian path-based method and formulates the preemption decision as an Integer Linear Programming (ILP) model. The model features a fixed number of variables, proving the problem to be Fixed-Parameter Tractable (FPT). A genetic algorithm is then applied to optimize task sequences and interference handling priorities. We evaluate our approach across 72 diverse scenarios (with 20 instances each), showing that it consistently outperforms existing methods in minimizing makespan. A secondary analysis on waiting times further confirms the operational benefits. To the best of our knowledge, this is the first study to address preemption decisions in TYCS under bidirectional flows. Our results demonstrate that the proposed method is both robust and scalable, with strong potential for improving real-world container terminal efficiency.

3 - A Preprocessing Framework for Narrowing Objective Domains in Challenging Integer Programs: Application to the Traveling Salesman Problem

Aadya Bhattarai, University of south florida, Tampa, FL, United States, Hadi Gard, LAGNAJITA BASU

We study how the set of possible optimal objective values in integer programs can be narrowed to a significantly smaller subset. To achieve this, we propose a branch-and-bound technique integrated with a constraint programming solver, applied to the binary-encoded reformulation of challenging integer programs. The primary goal of this approach is to significantly reduce the set of feasible optimal objective values, which in turn can lead to modified integer programming models that can potentially enhance the overall performance of solvers on difficult instances. We provide initial numerical results that support the efficacy of this approach. Specifically, we apply the proposed techniques to hard instances of the Traveling Salesman Problem and provide insights into its performance. Additionally, we demonstrate that the embedded bounding mechanisms within the proposed branch-and-bound framework, when warm-started, can significantly boost the performance of the overall approach.

4 - Stochastic Fleet Mix Optimization for Robot-Aided Intralogistics

Arghadeep Mitra, University of Missouri, Columbia, MO, United States, Sharan Srinivas

Intralogistics operations, a critical component of the supply chain, are among the most labor-intensive and time-consuming processes that directly impact overall efficiency and responsiveness. In recent years, the growing supply chain challenges, including labor shortages, cost constraints, rising turnover, and increased customer demands, have intensified the need for resilient solutions. The effective integration of autonomous mobile robots (AMRs) has the potential to address these challenges through collaborative automation, where AMRs handle labor-intensive order transport tasks, allowing human workers, who are adept at item retrieval, to focus on this more specialized task. While existing studies have investigated the operational impacts of AMRs in intralogistics, strategic decisions concerning optimal AMR fleet

composition remain unaddressed. This research proposes a two-stage stochastic model to minimize both operational and acquisition costs. The first stage addresses the strategic decisions regarding fleet mixing and sizing, while the second stage optimizes operational decisions, specifically order batching, batch assignment to AMRs, and sequencing. A sample average approximation approach is used to solve the model and estimate the expected total cost. Due to the NP-hard nature of the problem, a hybrid metaheuristic approach is developed to efficiently solve realistic instances. The applicability and effectiveness of the proposed method are demonstrated through extensive numerical analysis. In addition, several practical implications on AMR fleet sizing and mixing are also established based on the experiments.

Keywords: intralogistics, autonomous mobile robots, fleet mix, stochastic optimization, sample average approximation, metaheuristic

5 - A Column Generation Algorithm with Hybrid MILP+CP Decomposition for Irregular 3D Bin Packing in Maritime Block Transportation wonjun choi, Pusan National University, Pusan, Korea, Republic of

The Irregular-shaped Three-Dimensional Bin Packing Problem (I3DBPP) for maritime block transportation involves efficiently loading shipbuilding blocks into self-propelled barges while ensuring collision-free placement in discretized 3D space. The objective is to minimize the number of barges used, thereby reducing transportation cost and improving operational efficiency. This problem is NP-hard due to the combination of irregular 3D geometries, complex geometric constraints, and combinatorial assignment decisions. To address this challenge, we propose a column generation algorithm based on a hybrid MILP+CP decomposition. The master problem adopts a set covering formulation to select packing patterns, while the pricing problem generates feasible patterns via a two-stage subproblem: MILP-based 2D placement generation followed by CP-based 3D feasibility checking using no-overlap constraints. This hybrid approach leverages the modeling strengths of MILP and the constraint-handling capabilities of CP to efficiently tackle the spatial complexity of I3DBPP. Computational experiments on realistic shipyard datasets demonstrate the effectiveness of the proposed method in terms of solution quality and computational efficiency. This work contributes to advanced decision-support tools for shipbuilding logistics and highlights the potential of integrating MILP and CP in column generation frameworks.

6 - Fully Packed and Ready to GO: High-Density, Rearrangement-Free, Grid-Based Storage and Retrieval Tzvika Geft, Rutgers University, New Brunswick, NJ, United States, Kostas Bekris, Jingjin Yu

Grid-based storage systems with uniformly shaped loads (e.g., containers, pallets, totes) are commonplace in logistics, industrial, and transportation domains. A key performance metric for such systems is the maximization of space utilization, which requires some loads to be placed behind or below others, preventing direct access to them. Consequently, dense storage settings bring up the challenge of determining how to place loads while minimizing costly rearrangement efforts necessary during retrieval. This paper considers the setting involving an inbound phase, during which loads arrive, followed by an outbound phase, during which loads depart. The setting is prevalent in distribution centers, automated parking garages, and container ports. In both phases, minimizing the number of rearrangement actions results in more optimal (e.g., fast, energy-efficient, etc.) operations. In contrast to previous work focusing on stack-based systems, this effort examines the case where loads can be freely moved along the grid, e.g., by a mobile robot, expanding the range of possible motions. We establish that for a range of scenarios, such as having limited prior knowledge of the loads' arrival sequences or grids with a narrow opening, a (best possible) rearrangement-free solution always exists, including when the loads fill the grid to its capacity. In particular, when the sequences are fully known, we establish an intriguing characterization showing that rearrangement can always be avoided if and only if the open side of the grid (used to access the storage) is at least 3 cells wide. We further discuss useful practical implications of our solutions.

WA03

Building A Level 3 A303

Global Risk, Equity, and Policy Modeling

Contributed Session

Contributed

Chair: Sein Kim, Kyungpook National University, Daegu, Korea, Republic of

2 - Portfolio Selection Models with Weight Constraints Derived from a Nonlinear Shrinkage Method Reflecting Macroeconomic Variables: Macroaugmented Portfolio Models

Jihye Yang, Yonsei University, Seoul, Korea, Republic of, Seongmoon Kim

We propose portfolio selection models with weight constraints reflecting four important macroeconomic variables: the Baltic Dry Index, interest rates, money supply, and inflation rates. We design two types of macroaugmented portfolio models (MAPs): MAPs with a linear shrinking factor (MAPsL) and MAPs with a nonlinear shrinking factor (MAPsN). Furthermore, we incorporate four nonlinear shrinking factors into MAP construction with a combination level derived from a tracking signal (MAPcN). We assess the performance of MAPs compared to the benchmarks in 12 different stock markets, finding that MAPs outperforms the benchmarks. MAPsN outperforms MAPsL, and MAPcN outperforms MAPsN in terms of the Sharpe, Sortino, and Rachev ratios. Additionally, we compare the performance of MAPs. In terms of individual and collective measures, each MAPsN outperforms the corresponding constituent of MAPsL. The effectiveness of the nonlinear shrinkage method measured by the ratio of the Sharpe ratio (RS) is differentiated by the stock market's characteristics. The RS for each macroeconomic variable is closely related to trade dependence, financial market openness, money supply stability, and market concentration.

3 - Quantification of International Trade Network Performance Under Disruptions to Supply, Transportation, and Demand Capacity, and Exchange Rates in Disasters

Dana Hassani, University of Massachusetts Amherst, Amherst, MA, United States, Anna Nagurney, Oleg Nivievskiy, Pavlo Martyshev

Both sudden-onset and slow-onset disasters are causing disruptions to global trade, impacting the availability and affordability of commodities from agricultural to mineral ones. We develop a multicommodity international trade network equilibrium model under disaster scenarios with distinct probabilities of the occurrence of the disasters and their impacts on the capacities associated with production, transportation, and consumption. The disaster scenarios can also affect the exchange rates. We state the governing equilibrium conditions and derive the variational inequality formulation in commodity path flow variables and Lagrange multipliers associated with the capacity constraints. For each disaster scenario, we construct an international trade network performance measure, followed by a unified performance measure that includes all the disasters and their probabilities. Robustness is then quantified as the difference between the network performance under no disruptions and the unified performance

measure. An international trade network component performance indicator is also given to assess the impacts of the complete removal of trade network supply markets, demand markets, and/or transportation routes. The modeling framework is then illustrated through a series of numerical examples, motivated by Russia's war on Ukraine. The work is of relevance to decision-makers and policy-makers.

4 - Sustainable Administrative Management from the Perspective of Balanced Contraction

Yuji Sato, Chukyo University, Nagoya, Aichi, Japan

This study considered the state of sustainable administrative management from the perspective of balanced contraction of its scope of responsibility. Countries like Japan that have entered aged societies with low birthrate can no longer expect the increase of administrative resources. Therefore, efficient administrative management has become an urgent issue for local governments in Japan. This situation gives rise to the discussion of scaling back the current range of administrative services. However, the discussion is not straight forward due to vested interests of citizens or bureaucratic conflict within a government. Specifically, it is difficult to reach a consensus on the downsizing of services or the delegation of authority from a government to substitutable business operators. Public administration in such countries thus need to explore practical ways to show specific and logical grounds for the retrenchment. Theoretical approaches based on the New Public Management to this issue exist, while concrete ways of administration have not been presented to date. Focusing on the properties of publicness, this study proposed how to outsource public services to substitutable operators by evaluating the degree of matching between services and operators. In the evaluation, we applied the Analytic Hierarchy Process and the Principal Component Analysis to measure the properties of publicness and explored the combination of services and operators with the highest compatibility. A case study in a local government in Japan suggests the feasibility of the approach proposed in this study.

5 - Welfare vs. Revenue: Optimal Government Loan Design for Green Production

Minjeong An, Yonsei University, Seoul, Korea, Republic of, Jiyoung han, Seung Hwan Jung, Yunzhe Qiu

We explore government loan contract designs aimed at supporting green production while balancing environmental objectives and fiscal prudence. Traditionally, government finance has relied on subsidies that prioritize social welfare. However, increasing demand for substantial investments in green production has led to the emergence of government loans. Such loans uniquely combine aspects of traditional government subsidies—emphasizing social welfare—and supply chain finance—prioritizing revenue maximization and risk mitigation. Employing a Stackelberg game framework, this study systematically compares these two distinct policy approaches: social welfare maximization versus government revenue maximization subject to minimum green production constraints. Our analytical model demonstrates significant divergence in optimal interest rate policies and production outcomes under these contrasting objectives. Specifically, welfare-oriented governments typically adopt zero interest rates to enhance environmental outputs, whereas revenue-oriented policies impose positive interest rates to mitigate fiscal risks and ensure repayment. Furthermore, we introduce an innovative two-part financing contract capable of reconciling these conflicting goals. This contract structure decouples marginal borrowing costs from fixed fees, enabling governments to achieve socially optimal production levels without necessitating negative interest rates. Our findings contribute valuable insights for policymakers designing green loan programs, highlighting essential trade-offs between ecological benefits and fiscal accountability in sustainable financing initiatives.

6 - Linking Global Supply Pressures to Business Cycle Dynamics: a Framework for Enhancing the Gscpi

Sein Kim, School of Business Administration, Kyungpook National University, Daegu, Korea, Republic of, Sungsu Kim

The Global Supply Chain Pressure Index (GSCPI), developed by the Federal Reserve Bank of New York, provides a valuable benchmark for quantifying global supply disruptions. However, its current formulation faces limitations in capturing the evolving, nonlinear nature of supply shocks. In particular, the index does not incorporate commodity price movements, which play a critical role in signaling upstream supply stress and in the transmission of input costs across production networks. Moreover, its reliance on static principal component analysis (PCA) limits the model's responsiveness to structural shifts and time-varying transmission mechanisms. This study explores the development of an enhanced GSCPI that addresses these limitations by incorporating additional price-based indicators and adopting a dynamic PCA framework. To evaluate its macroeconomic relevance, we apply standard business cycle moment diagnostics volatility, cyclical with output, persistence, and lead/lag relationships widely used in empirical validation of macroeconomic models. By bridging high-frequency supply-side indicators with low-frequency macroeconomic fluctuations, this approach contributes to a more adaptive and interpretable assessment of global supply pressures and offers a framework for more timely macroeconomic assessment.

WA04

Building A Level 3 A304

Climate Risk, Adaptation, and Energy Policy

Contributed Session

Contributed

Chair: Raziye Aghapour, University of Texas at Arlington, Arlington, TX, United States

1 - Paying for Drawdown: the Value of Commitment on the Cost of Ending Global Warming

John F Raffensperger, Simon Business School, University of Rochester, NY, Rochester, NY, United States

This paper investigates the costs of ending global warming under different levels of commitment. To estimate the costs, the paper contributes “smart market” models simulating auctions for emissions and carbon removal to reach net zero or base zero (pre-industrial temperature) by a particular date.

The models constrain warming robustly under different discount rates. The paper shows that Hotelling's rule does not apply to carbon policy with a deadline.

The paper contributes a mechanism to pay for drawdown of warming beyond net zero, avoiding third party payments. The mechanism converts a revenue negative auction to revenue neutral. The paper shows that drawdown costs less when emitters pay for it than when a third party pays for it.

The paper contributes a quantitative method to model and price carbon removal contracts, and gives estimates of the value of commitment to strong versus weak contracts.

The least cost trajectory requires long commitments to emissions reductions and carbon removal. The paper gives estimates of the value of long-run versus short-run commitments and the value of the ability to manage revenue across decades.

The paper provides a method to calibrate the models to a climate simulator while accounting for all greenhouse gasses.

For reducing global temperature 1.4°C by 2125, estimates range from \$20.4 trillion to \$10.8 trillion (total over 100 years present value at 3%) depending on commitment. This is much lower than other researchers' estimates simply for keeping global temperature to 1.5°C. The models could be used in actual trading.

2 - Climate and Water Risk Across the Bulk Power System Under Various Future Climate and Infrastructure Buildout Scenarios

Jonathan Lala, EPRI, Palo Alto, CA, United States, Erik Smith, Nalini Rao

The bulk power system in the U.S. faces substantial changes in the coming decades stemming from both climate change and the shift of the generation mix towards renewables. These changes present a challenge for both system planners and operators, who need a solid understanding of how changes will manifest as risks across uncertain climates and system configurations. Here we present a standardized and consistent process to identify relevant climate hazard metrics, connect them to vulnerabilities, and measure impacts across thermoelectric, hydropower, wind, solar, and transmission assets. We then simulate this process across various future climate models, emissions scenarios, and power system infrastructure buildouts to characterize uncertainty across the risk space. Results demonstrate unique regional risk profiles

depending on the climate hazard, climate scenario, and generation mix, although risks from high temperatures are measurable across all assets and regions. Stakeholders can use this risk-based approach to identify and monitor the effects of capacity losses and potential resulting consequences as climate and infrastructure change.

3 - Environmental Flows and the Hidden Costs of Hydropower: Evidence from Glen Canyon Dam

Ao Yu, Argonne National Laboratory, Lemont, IL, United States, Matija Pavicevic, Quentin Ploussard

Hydropower production in the U.S. operates within a complex regulatory framework that balances power system reliability, resource allocation, and environmental regulation. Because power plants are interconnected through the grid, policies affecting hydropower can have spillover effects on the broader generation fleet. This paper evaluates the economic impact of a 2024 environmental flow experiment that required dam operators to bypass water rather than use it for electricity generation. Leveraging quasi-random variation in the bypass schedule and an econometric model of hourly unit-level generation data, we estimate that the experiment increased system-level generation costs substantially. The implied economic value of the foregone water was about \$20 to \$50 per acre-foot. In addition, the policy resulted in significant environmental costs in increased pollutant emissions due to substitution toward fossil-fuel generation. These findings have important implications for designing environmental flow regulations and electricity market policy, particularly in the context of growing water scarcity and increased variable renewable energy penetration.

4 - Designing a Resilient Renewable Energy System for Zero-Energy Buildings Using a Simulation-Based Stochastic Program

Ricardo Lima, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, Farah Souayfane, Raju Pathak, Hayssam Dahrouj, Ibrahim Hoteit, Omar Knio

This work presents a methodology for designing resilient renewable energy systems for zero-energy buildings that accounts for the uncertainty of climate change over the next 75 years. Zero-energy buildings are poised to be energy-efficient and feature on-site generation and storage to meet their power demand. It becomes, therefore, a fundamental quest to properly design a resilient renewable energy system and storage capacities to meet the building's electricity demand while minimizing the life-cycle cost. The proposed methodology consists of 1) generation of long-term high-resolution atmospheric data by downscaling global climate projections, 2) a building simulation to estimate cooling electricity demand as a function of the atmospheric input fields, 3) a renewable energy system simulation integrated alongside the load-demand forecast, 4) a stochastic program based on climate projection scenarios; 5) adaptation of the stochastic programming for implementation within a simulation-based optimization framework, 6) a multi-objective optimization using a simulation-based algorithm to determine the best system configuration balancing resilience and cost, and 7) a rolling-horizon approach applying the reformulated stochastic program over 25-year horizons at hourly resolution, and 8) integration of extreme weather events from multiple climate models into the optimization framework to ensure robustness. Our optimization results emphasize the first-stage decision variables, and their associated 25-year expected life cycle cost of the renewable energy system. In particular, we quantify how incorporating extreme weather events from multiple climate model scenarios alters both the system design and its life-cycle cost.

5 - A Design of Experiments Approach to Assessing State-Level Decarbonization Strategies in Texas Using GCAM-Usa

Raziye Aghapour, University of Texas at Arlington, Arlington, TX, United States, Erick Jones

Effective state-level decarbonization planning requires an understanding of how different policy levers interact across the energy, industrial, and transportation sectors. This study presents a structured Design of Experiments (DoE) approach for systematically evaluating key policy pathways using GCAM-USA, a state-resolved integrated assessment model. The analysis focuses on Texas as a case study and investigates the influence of five continuous policy variables: adjustments to natural gas prices, subsidies for rooftop solar adoption, constraints on electric vehicle (EV) market share, production caps in the cement industry, and taxation of oil refining activities.

To efficiently explore the multidimensional policy space, we use a **Sobol sequence**—a quasi-random low-discrepancy sampling method—to generate a balanced and representative set of 35 scenarios. These scenarios are designed to capture a wide range of potential interactions between policy levers. Total CO₂ emissions are tracked as the primary output to assess how policy combinations influence mitigation outcomes.

This work highlights the value of combining integrated assessment modeling with structured experimental design to support subnational climate policy evaluation. By enabling transparent, reproducible, and systematic scenario analysis, this approach helps identify promising decarbonization strategies that account for sectoral interdependencies and policy trade-offs.

WA05

Building A Level 3 A305

Toward Circular and Low-Carbon Transitions: Policies, Incentives, and System Designs

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Ilgin Dogan, INSEAD, Fontainebleau, France

1 - Closing the Loop OR Opening the Market? Deposit-Refund System Design Under Recycling Targets

Donghyun Choi, Georgia Institute of Technology, Atlanta, GA, United States, Andre Calmon, Beril Toktay

Beverage producers increasingly face mandatory recycled-content targets while financing deposit-refund systems (DRS) or broader extended-producer-responsibility schemes. Yet DRS governance differs markedly, reshaping both system costs and material circularity. We build an analytical model that compares (i) open-access, in which DRS-collected material is sold to the competitive commodity market, with (ii) priority-access, in which beverage producers have the right of first refusal (ROFR) on that material. The model embeds supply uncertainty and competition from non-beverage sectors to quantify how each design shifts financial burdens, price risk, and closed-loop (bottle-to-bottle) recycling rates. By illuminating the trade-offs, the study provides actionable guidance for policymakers crafting DRS legislation and for beverage firms shaping sourcing and investment strategies in a circular-economy context.

2 - Diffusion in Circularity: the Missing Link in Circular System Design

Clara Carrera, INSEAD, Fontainebleau, France, Atalay Atas, Dragos Florin Ciocan, Luk Van Wassenhove

Despite growing enthusiasm for circular economy models, implementing profitable circular strategies remains challenging—particularly for new products with short lifecycles and unpredictable return flows. This paper examines the viability of circular business models that combine leasing (servitization) with remanufacturing, focusing on how product diffusion dynamics shape the alignment between return supply and remanufactured demand. Using a dynamic model that captures consumer heterogeneity, arrival patterns, and word-of-mouth effects, the analysis introduces the concept of Circular Potential—a measure of the temporal overlap between returned products and remanufactured demand. The results show that products with slower diffusion and stronger peer influence are better suited for circular models, while products targeting environmentally conscious consumers often exhibit lower Circular Potential due to faster uptake and longer usage times. The study identifies conditions under which managers can ignore or must account for lifecycle effects in pricing and contract design. Calibration with data from the consumer electronics industry highlights the counterintuitive finding that mass-market products may offer more favorable conditions for circularity than sustainability-focused alternatives.

3 - An Empirical Analysis of Sustainable EV Charging: Human vs. Smart Algorithm

Rahul Roy, Kenan-Flagler Business School, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Nur Sunar, Jayashankar Swaminathan

The transportation sector is the world's second-largest source of greenhouse gas (GHG) emissions, making its decarbonization essential for global climate goals. Electric vehicles (EVs) offer a promising pathway to sustainable mobility by reducing fossil fuel dependence. However, widespread EV adoption increases electricity demand on

the grid, especially during peak consumption hours, when electricity is costlier and harder to supply. This creates a challenge for both grid operators and consumers. Shifting EV charging to off-peak hours may reduce system strain and lower user costs, benefiting all stakeholders. As such, utilities and grid operators have adopted various charging strategies to foster off-peak charging: some allow fully consumer-controlled charging, while others rely on automated algorithms or hybrid approaches that incorporate both algorithmic control and user feedback. In this context, we study two key questions: (i) How do smart algorithms perform relative to humans in promoting off-peak EV charging? (ii) Building on this comparison, how does incorporating human guidance into smart algorithmic systems affect off-peak charging performance? To answer these questions, we analyze data from a large-scale field experiment on residential EV charging. Our findings offer valuable insights and practical guidelines for grid operators and utilities designing smart charging systems, especially those involving human-algorithm interaction.

4 - Area Conditions and Positive Incentives: Engaging Local Communities to Protect Forests

Xavier Warnes, Indiana University, Bloomington, IN, United States, Joann de Zegher, Dan Iancu, Erica Plambeck

Agricultural production is a dominant driver of tropical deforestation, causing alarming CO₂ emissions and loss of biodiversity and ecosystem services. Complex land tenure systems coupled with low enforcement of conservation laws enable farmers to convert tropical forests into productive land for a better income. We propose positive incentives with area conditions for local communities to prevent this deforestation while increasing farmer welfare. We model the deforestation decisions of the community using a cooperative game in partition correspondence form and show how these area-conditional incentives can leverage the existing social capital in agricultural communities. Finally, we demonstrate the effectiveness of these conditions by using data from the Indonesian context.

WA06

Building A Level 3 A306

Coordination in E-Commerce Supply Chains

Contributed Session

Contributed

Chair: Rafael Becerril Arreola, USC Darla Moore School of Business, Columbia, SC, United States

1 - Competing Platforms, Competing Sellers: the Strategic Impact of Personalized Recommendations in Two-Tiered Market Competition

Zhe Yang, Tianjin University, Tianjin, China, People's Republic of, Zhaofang Mao, Yuxin Liu, Yunchuan Liu

Personalized recommendations, powered by advanced machine learning techniques, have become a key strategic tool for platforms. We explore the design choices between profit-based and demand-based recommendation systems when competing sellers offer products of varying quality to heterogeneous consumers through online platforms. Our analysis considers both monopolistic and competitive platform settings. Our findings reveal the following: (1) In a monopolistic market, the platform consistently prefers profit-based recommendation systems. However, in a duopoly, a platform may adopt a demand-based system to alleviate the double-layered competition between platforms and sellers. (2) In a duopoly, the pioneering platform always selects a profit-based recommendation system, regardless of the product cost structure or the degree of consumer heterogeneity. (3) In contrast, as consumer heterogeneity increases, the following platform becomes more inclined to adopt a demand-based recommendation system. (4) A profit-based recommendation system does not necessarily promote high-quality products; the recommendation outcome depends on the seller's cost structure and consumer heterogeneity. Our paper also offers managerial implications for policymakers regarding the regulation of recommendation algorithms and the protection of consumer privacy.

2 - Information Disclosure and Efficiency of Dynamic Price Negotiations

Gaoyang Cai, Northwestern University, Kellogg School of Management, Evanston, IL, United States

Negotiation is prevalent in two-sided markets, where users have incomplete information about the other side. Information disclosed by the platform owner or revealed by users critically impacts market outcomes and welfare.

Leveraging a panel experiment on a large on-demand delivery platform that exogenously varies the information disclosed to drivers negotiating with shippers on prices, I find that disclosing information about the lowest counteroffer rejected by a shipper significantly increases shipper rejections on driver counter prices while increasing transactions due to the decreasing driver counter prices.

To evaluate the equilibrium and welfare impacts of different information disclosure policies, I develop a dynamic model incorporating shipper counteroffer acceptance and driver counter strategies under different information sets. I find that disclosing the information benefits low-value shippers by allowing them to signal valuations and decrease subsequent counter prices by rejecting counteroffers and waiting. In contrast, high-value shippers suffer from drivers' more aggressive countering based on their rejection records.

Counterfactual analyses show that a partial information disclosure policy, i.e., no information disclosure above a threshold and full information disclosure below the threshold, protects high-value shippers from price discrimination while keeping the signaling benefits. The platform can restore 91.45% of the social surplus in the infeasible first-best benchmark without incomplete information and driver markups by optimally choosing the partial information disclosure threshold. The optimal partial disclosure policy increases the social welfare by 19.96% compared to the no-information-disclosure benchmark and increases the social welfare by 11.25% compared to the full-information-disclosure benchmark.

3 - Multi-Tier Capacity Planning in Fulfillment Center and Storage Networks at Amazon

Junho Lee, Amazon, New York, NY, United States, Song Song

From the business problem, we develop a science model answering when, how many, where, and what to build, with uncertainty in future demand and economic factors coupled with capacity planning. It supports capacity planning of the fulfillment center (FC) and storage and distribution centers (SDCs) at Amazon. We introduce an approach that decomposes a monolithic model into two parts: topology and economics models. Given how many to build, the topology model generates a build plan, which consists of the best locations of FCs and SDCs. The economics model evaluates economic factors for every build plan from the topology model and recommends the one with the least cost. Our model has been supporting multi-billion dollars' decision making for Amazon's FC and SDC network.

4 - The Steady State of Picking and Packing Throughputs in Fulfillment Centers, with The Optimal Labor Allocation

Jiung Lee, Coupang, Seoul, Korea, Republic of, Zhixin Liu, Ryan Cybulski, Vikram Balaji, Manish Garg, Gowri Krishnan, Paul Stroup, Harshal Wanjari

In the domain of e-commerce fulfillment, optimizing labor allocation within fulfillment centers is critical to ensuring high throughput and meeting stringent service level agreements. This study introduces a novel analytical model that captures the end-to-end dynamics of picking and packing operations for fresh products—a process characterized by multiple SKUs in an order, heterogeneous labor resources, and tight operational constraints. Our model explicitly incorporates the interplay between pickers, item collection walls, and packing stations, all of which must be coordinated to maintain flow balance and avoid bottlenecks.

We develop a mathematical framework that characterizes the steady-state throughput capacity of each resource type and identifies optimal allocation strategies under realistic operational constraints. The model enables decision-makers to evaluate trade-offs and reconfigure resources dynamically to maximize efficiency.

To validate our approach, we calibrate the model using proprietary operational data from a global commerce service provider. Empirical results demonstrate the model's ability to accurately capture system behavior and provide actionable insights for improving fulfillment performance. Our findings offer both theoretical contributions to the modeling of complex logistics systems and practical implications for real-world deployment in high-volume, time-sensitive environments.

5 - Online Supplier Selection for a Drop Shipping E-Retailer

Thomas Bryant Cassidey, University of Colorado Colorado Springs, Denver, CO, United States, Emily Griffin, Burcu Keskin

Firms using drop shipping must choose suppliers to fulfill orders in real time. Suppliers may vary in fulfillment accuracy, fulfillment speed, order acceptance rate, and cost, among other attributes. These trade-offs are difficult to manage due to the difficulty in eliciting coherent preferences from decision makers and due to the uncertainty of important supplier attributes. Further, due to the on-line nature of the problem and suppliers both leaving and joining the network, varying levels of information are available to a firm when an order is received. We devise a system to elicit the preferences of a firm and investigate methods to automate decisions accordingly, considering known factors of the supply options and a prediction of their performance. We test our methods using real transactional data from a large e-retailer of durable goods in the United States. Our case study provides an analysis of the important factors in predicting supplier attributes and demonstrates the trade-offs in outcomes of our methods compared to the incumbent approach.

6 - Auditing the Effectiveness and Impartiality of Pricing Algorithms in Online Grocery Retailing

Rafael Becerril Arreola, USC Darla Moore School of Business, Columbia, SC, United States

Online retailers often use automated software to set product prices that maximize profits by capitalizing on geographic differences of consumer sensitivity to prices. As a result, prices may correlate with consumer demographics. We collect and analyze data to study the variation of algorithmic prices across geographies and to investigate what explains such geographic differences. We compare online prices against brick-and-mortar prices to isolate the algorithmic component of prices and find it to be correlated with some demographics. By reverse-engineering the pricing algorithm, we find that it is not fully consistent with profit maximization. The data shows that the algorithm is suboptimal for products and locations with low product sales, suggesting that the algorithm is ineffective because of weaknesses in the data it uses. As a result, the pricing algorithm can only effectively maximize profits for consumers with specific demographic profiles.

WA07

Building A Level 3 A307

Strategic Innovation and the Future of Technology Adoption

Contributed Session

Contributed

Chair: Junyong Lee, 1732, Deogyong-daero, Giheung-gu, Yongin-si, Gyeonggi-do, Republic of Korea, Yongin-si, Gyeonggi-do., 17104, Korea, Republic of

1 - Selecting Industry 4.0 Technologies for Applications in Digital Construction Using Text Mining and Clustering Techniques

Hamid Seifoddini, University of Wisconsin-Milwaukee, Milwaukee, WI, United States, Benyamin Sadeghi

Industry 4.0 technologies—such as Artificial Intelligence (AI), robotics, 3D printing, and the Internet of Things (IoT) are essential for enhancing the efficiency and effectiveness of human endeavors in the 21st century. This is especially relevant in the construction industry, which has lagged manufacturing in adopting these advanced technologies. To address this gap, we have developed a comprehensive framework for identifying and classifying 33 Industry 4.0 technologies based on 106 application attributes specific to construction, using text mining and clustering techniques.

In this paper, we introduce a framework for modeling the alignment of advanced technologies with various construction activities using clustering techniques. To prioritize these technologies, we assign weighting factors to key attributes such as safety, productivity, sustainability, cost, quality, and timeliness. This approach allows users to tailor solutions according to their specific priorities. Moreover, by applying a range of weighting factors, practitioners can conduct sensitivity analyses to identify the most suitable cluster of advanced technologies for a given priority. This study also highlights the power of data mining in transforming large volumes of data into concise, useful, and actionable information.

The framework is grounded in an extensive review of over 500 peer-reviewed publications and additional resources. Its application supports the advancement of Construction 4.0 by promoting the integration of Industry 4.0 technologies into construction practices.

2 - From Experimentation to Impact: Governing Enterprise Innovation Through Crucial use Case Identification

Joseph Ng, NYU, New York, NY, United States, Andres Fortino

Organizations face a critical challenge in translating burgeoning emerging technology experiments into scalable, impactful enterprise solutions. This innovation bottleneck stems from a "portfolio fallacy," prioritizing exploration volume over rigorous, strategically aligned evaluation. Drawing upon the scientific concept of *experimentum crucis* and the "killer uses" framework, we posit that identifying crucial use cases—high-leverage business needs uniquely addressable by new technologies—is paramount for overcoming this challenge. Unlike retrospective "killer apps," crucial use cases are forward-looking and amenable to systematic identification and evaluation.

This paper introduces the Casus Exponentialis Technology Matrix (CETM), a multi-dimensional governance framework designed to operationalize the analytical identification and evaluation of crucial use cases. Leveraging principles from decision theory and innovation diffusion, CETM provides a structured, quantitative scoring mechanism across eight critical dimensions: technology readiness, business impact, ethical implications, regulatory compliance, cybersecurity, workforce impact, sustainability, and market risk. By applying this systematic evaluation, CIOs and innovation leaders can move beyond intuitive decision-making, prioritize initiatives with demonstrable potential for responsible scaling, and align technology investments with strategic objectives. The CETM framework fosters interdisciplinary accountability and supports a repeatable innovation lifecycle from initial discovery through institutionalization. This approach offers a robust methodology rooted in management science for navigating the complexities of the exponential technology landscape, transforming fragmented experimentation into structured, value-driven transformation, and directly contributing to the revolutionizing of decision-making through the effective integration of technology and human collaboration.

3 - A Strategic Framework for Identifying National R&D Investment Opportunities in Energy Transition: An Integrated Analysis of Patent, SCIE, and SSCI Data

Heeseok Moon, Kyung Hee University, Yongin, Korea, Republic of, Hyunhong Choi

This study presents a strategic framework for identifying national R&D investment opportunities in the energy transition sector by integrating patent data with SCIE and SSCI publication records. To address the limitations of existing technology-centered analyses, the study incorporates signals from scientific discovery and societal discourse.

A five-stage analytical process is proposed. First, Structural Topic Modeling (STM) is applied to extract latent topics from patents and academic papers, integrating metadata to estimate context-sensitive topic prevalence. Second, regression analysis is performed to evaluate topic-level characteristics such as growth, novelty, and citation performance, enabling the identification of emerging topics. Correlation analysis is then used to cluster closely related topics, capturing recent trends in technological convergence. Third, the analysis assesses national R&D focus by examining country-specific topic intensities and regression coefficients. Fourth, TF-IDF-based cosine similarity is calculated to evaluate semantic alignment between patents and papers, identifying both shared research-industry trends and academically dominant topics. Fifth, a set of decision-support tools is developed: (1) a Patent Portfolio Matrix that maps technology-level impact and country-level competitiveness, and (2) a commercialization opportunity map derived from emerging academic topics that are weakly aligned with patents but closely related to societal concerns.

By linking scientific, technological, and social dimensions of innovation, this framework offers a data-driven basis for forward-looking R&D strategy development. It supports the identification of underexplored yet impactful investment areas and contributes to national-level planning in complex, demand-sensitive domains such as

energy transition.

4 - Pricing Problems in Adoption of New Technologies

Yijin Wang, University of Illinois Urbana-Champaign, Urbana, IL, United States, Subhonmesh Bose

We propose a generalization of the Bass diffusion model in discrete-time that models the effect of price in adoption. Our model fits well to adoption data for various products and is amenable to mathematical analysis. We utilize this model to study two decision-making problems. First, we provide a series of structural results on optimal pricing strategies to maximize profits from product sales by a monopolist over a finite horizon. We derive a closed-form optimal pricing strategy for the single-period problem. For the multi-period problem, we show that the search for the optimal price can be restricted to a bounded set. Furthermore, for the special case of low peer effects, we analytically establish some desirable properties of the value function and the optimal price function. We also show that the optimal pricing trajectories calculated using our model have a similar structure to those proposed by prior research. Second, we study a Stackelberg game between a policy-maker and a monopolist, where the former seeks to maximize adoption through rebates, while the latter maximizes profits. The firm, which is the follower, must determine its optimal pricing strategy in response to a fixed rebate set by the leader, the policymaker. We analytically characterize the single-period game and show that the presence of rebates does indeed boost adoption. Due to the difficulty in analyzing the multi-period game, we conduct numerical simulations and show that characteristics of the single-period game appear to apply to the multi-period game as well.

5 - Cultivating Future Innovators: a Smart Agriculture Summer Camp for Middle and High School Students

Sara Masoud, Wayne State University, Detroit, MI, United States

This presentation highlights the objectives, structure, and projected outcomes of the Smart Agriculture Summer Camp. Funded by USDA and supported by additional partners, the camp aims to introduce middle school and high school students to the emerging field of smart agriculture through immersive, hands-on learning experiences. Over the two-week period, students are engaged with four interconnected modules: Indoor/Outdoor Farming, Internet of Things (IoT) and Sensor Technologies, Data Analysis, and Artificial Intelligence/Machine Learning in Agriculture. Each module combines experiential learning activities, expert-led workshops, and site visits to local farms and tech labs, offering students a well-rounded understanding of the technologies transforming agriculture. The program is intentionally designed to bridge the gap between traditional agricultural practices and advanced digital tools, while fostering awareness of STEM careers in agri-tech. To measure the program's impact, pre- and post-camp surveys are administered to assess changes in students' knowledge, attitudes, and interest in agricultural technologies and related career pathways. The camp also incorporates leadership development and peer collaboration exercises to build students' confidence and problem-solving skills.

6 - Technology Opportunity Discovery for Incremental Innovation: A Hybrid Framework with Structural Topic Modeling and Generative Topographic Mapping

Junyong Lee, Kyung Hee University, Yongin-si, Gyeonggi-do, Korea, Republic of, Hyunhong Choi

As markets are increasingly shaped by diverse technologies, the identification of technological opportunities has become paramount. This study proposes a novel framework combining Structural Topic Modeling and Generative Topographic Mapping to identify incremental innovation opportunities. While recent analyses of technological opportunities have predominantly focused on identifying radical innovations, the identification of incremental innovations holds comparable significance, as their accumulation can pave the way for radical breakthroughs.

The proposed framework uses patent data to identify latent topics, then constructs data points from principal topic keywords, reducing these to latent points. Using GTM's inverse function, topic models enhance the interpretability of vacant coordinates, offering a robust decision-support tool for unearthing technological opportunities. This approach adopts a semi-automated methodology, minimizing expert intervention during the analysis phase.

An empirical study was conducted on Unmanned Aerial Vehicles (UAVs), analyzing the trend of how technological gaps are filled over time according to patent application dates. From the technology map, 11 vacant points were identified and proposed as promising candidates for incremental innovation. Consequently, this research contributes to the discovery of incremental innovation opportunities by demonstrating that topic models can be employed not only to improve GTM's dimensionality reduction performance but also to simultaneously strengthen its interpretative capabilities.

Furthermore, the effectiveness of the proposed methodology was validated. This was achieved by inputting the generated outputs as prompts to ChatGPT and comparing the results obtained when utilizing comprehensive information from both STM and GTM against those derived solely from GTM-restored keyword information.

WA08

Building A Level 3 A310

Health IT and User Engagement: Patients and Clinicians

Invited Session

Information Systems

Chair: Atiye Cansu Erol, Johns Hopkins University-Carey Business School, Baltimore, United States

1 - Does Premium Version Adoption in Mhealth Improve User Engagement and Health-Related Outcomes?

Yikun Jiang, Purdue University Mitch Daniels School of Business, West Lafayette, IN, United States, Kosuke Uetake, Nathan Yang

This study examines the impact of premium adoption on user engagement and health outcomes in an mHealth setting. Using a staggered Difference-in-Differences (DiD) framework, we estimate the effects of premium adoption on food and exercise tracking, adherence to caloric budgets, and weight loss. Although premium users exhibit an initial spike in engagement, these effects dissipate within weeks, consistent with hedonic decline. Despite increased tracking activity, premium adoption does not translate into sustained improvements in weight loss. We further demonstrate that failing to account for endogeneity and selection in premium adoption leads to an overstatement of its effects on both engagement and health outcomes. Taken together, these findings underscore the transient nature of premium-induced behavior change and highlight the importance of reinforcement mechanisms to sustain long-term engagement.

2 - The Impact of Telehealth Expansion on 'Doctor-Shopping' and Drug Overdose

Byoung-Hyuk Ahn, Boston University, Boston, MA, United States, Meizi Zhou, Gordon Burtch

Telehealth technologies have revolutionized the traditionally fragmented and polarized U.S. healthcare landscape by enabling efficient and cost-effective access to quality healthcare services across state borders. Legislative initiatives at both state and federal levels such as Interstate Medical Licensure Compact (IMLC) and temporary licensure waivers during the COVID-19 pandemic further democratized healthcare access and accelerated telehealth adoption. However, this expansion has simultaneously introduced new pathways for patients to engage in "doctor-shopping" for controlled substances. In this study, we examine the causal relationship between telehealth expansion and patients' attempts in acquiring prescription drugs from multiple providers and related drug overdoses, leveraging the staggered timing of states' entry into the IMLC and national temporary licensure waivers in U.S. as a quasi-natural experiment. Analyzing the patient data from 2012 to 2022, our findings reveal that telehealth expansion increases doctor-shopping activities for controlled substances and associated overdose incidents, with nuanced variations across drug categories. Additionally, we investigate the moderating effects of Prescription Drug Monitoring Programs (PDMPs) and Interstate Data Sharing initiatives, highlighting the monitoring effect in mitigating these adverse outcomes. Finally, we investigate the cross-drug effects by analyzing how change in illicit drug price influences doctor-shopping behavior and overdose rates. Our study illuminates the unintended consequences of telehealth technologies and provides practical regulatory implications for healthcare system designers and policymakers.

3 - The Devil's Advocate: when AI Challenges Human Expertise

Kanghyun Cho, Temple University, Philadelphia, PA, United States, Kihwan Nam, Jin-Woo Park

While artificial intelligence (AI) has the potential to enhance clinical decision-making, much existing research focuses on low-stakes tasks with non-experts, overlooking the complexities of high-stakes environments where decisions involve expert judgment and serious consequences. To address this gap, we examine AI use in a high-stakes medical context through the lens of reflective practice. We conducted a randomized crossover experiment with experienced psychiatrists diagnosing swallowing disorders using video data. Two AI-assisted workflows were tested: a one-step workflow, where AI advice is given during initial diagnosis, and a two-step workflow, where AI input is revealed only after an initial unaided decision. The one-step workflow improved overall accuracy but increased automation bias, particularly in simple cases. The two-step approach improved accuracy for medium-difficulty tasks by promoting reflective engagement but decreased performance in highly difficult cases, likely due to cognitive conflict when AI contradicted prior assessments. Notably, our AI system outperformed human experts in many cases, but both reached a performance ceiling on the most challenging tasks. These findings underscore how expert identity, cognitive complexity, and the consequences of error influence AI effectiveness in clinical settings. We argue that results from low-stakes studies may not generalize to high-stakes environments and call for the development of task-sensitive, context-aware AI systems that complement rather than replace professional judgment.

4 - Offline Privacy Concerns in Mental Health Care Delivery Through Telehealth

Atiye Cansu Erol, Johns Hopkins University - Carey Business School, Baltimore, MD, United States, Lynn Wu

Mental health conditions are responsible for 23% of all pregnancy-related deaths in the US. A significant contributor is postpartum (after birth) mental health issues, primarily depression, which affect 1 in 8 mothers. Yet more than half of new mothers with depression go untreated. Telehealth has the potential to increase access to and utilization of mental health care, but any such benefits may be affected by offline(physical) privacy concerns. On the one hand, telehealth shifts care from a provider office setting to a home setting, which may reduce privacy for individuals using telehealth in crowded households(“privacy at home”). On the other hand, consumption of telehealth services may be less observable than in-person care, alleviating concerns about societal stigma related to the use of mental health services(“privacy in the community”). Using detailed insurance claims data surrounding a period of rapid increase in telehealth use, we find that telehealth availability increases treatment utilization, with lower utilization for individuals in crowded households and higher utilization where there is a high societal stigma to the use of mental services. These insights are potentially actionable as insurers and providers can expand the utilization of maternal mental health services by increasing support for telehealth in environments where there is high societal stigma, and seek alternative delivery mechanisms for individuals who may live in crowded households. Furthermore, our findings are likely to extend beyond maternal mental health setting, as the advantages and disadvantages of telehealth and the associated privacy concerns are not exclusive to mothers or mental health.

WA09

Building A Level 3 A311

People-Centric Operations

Invited Session

Manufacturing and Service Operations Management (MSOM)

Chair: Abhishek Deshmene, Georgia Institute of Technology, Atlanta, GA, United States

1 - Fair Funnels: Bias, Performance, and Interventions in Multistage Hiring Processes

Jordan Tong, University of Wisconsin-Madison, Madison, WI, United States, Eunjee Kim, Xiaoyang Long

How does the multistage funnel structure interact with discriminatory judgment behaviors to impact hire representation and quality? What is the best way to design regulations and incentives within funnel structures to improve representation and performance, and can they improve both? We develop a stylized mathematical model that captures tasted-based and statistical discrimination in a multistage funnel. We show that three distinct discriminatory judgment behaviors work through the funnel in three distinct ways to cause underrepresentation. We then endogenize the funnel's threshold decisions to study the impact and effective design of regulations and incentives. Whereas common wisdom would suggest targeting regulations and incentives at the stage where underrepresentation appears most severe, our analysis suggests a simple recommendation is surprisingly robust irrespective of where the underrepresentation manifests: *regulate the top, subsidize the bottom*. We identify several distinct asymmetric dynamics that support this prescription: asymmetry in regulation workarounds, asymmetry in regulation win-win opportunity, and asymmetry in subsidy propagation. Our results help organizations better predict how discriminatory judgment behaviors will affect representation and quality at various stages in the funnel. They help guide policymakers seeking to design regulations and incentives on where to target their interventions, and their likely consequences on representation and performance outcomes.

2 - Information Sharing to Optimize the Wait-TIME Experience

Laurens Debo, Dartmouth College, Hanover, NH, United States, Jeunghyun Kim, Robert Shumsky

We develop a behavioral model to optimize the frequency of queue-position updates provided to customers in an unobservable M/G/1 queue. Departing from standard waiting cost functions, the model captures disutility using two psychological constructs: *loss aversion* and *discomfort with uncertainty*. We evaluate a class of information policies in which customers receive position updates every $\$m\$$ service completions.

We first examine two extremes: a full-information policy ($\$m = 1\$$) and an arrival-only policy ($\$m = \infty\$$). Customers who are purely loss-averse prefer minimal information, while those who dislike uncertainty benefit from frequent updates. For customers sensitive to both, we derive a tractable approximation for highly utilized systems to determine the optimal update interval $\$m\$$. The analysis shows that as utilization or service-time variability increases, the optimal strategy is to reduce the frequency of updates (i.e., increase $\$m\$$). However, the *number* of updates experienced by the average customer still rises, as longer queues compensate for the sparser update schedule.

Numerical experiments validate the approximation and demonstrate robustness even under moderate utilization.

This work contributes a behavioral optimization framework to the queueing literature and provides actionable insights for service providers: tailoring the timing of updates can improve customer experience by managing emotional responses to waiting, not just expected wait times.

3 - Dali OR Dall-E? Consumer Perception of Human Versus AI Creative Processes

Abhishek Deshmene, Georgia Institute of Technology, Atlanta, GA, United States, Diwas KC

As AI-generated content proliferates in creative domains, understanding how consumers evaluate the authenticity and value of such works has become increasingly important. This study examines how attribution — human versus AI-assisted — shapes consumer perceptions of visual art, and whether disclosures about creative effort and uniqueness moderate these effects. Drawing on theories of authenticity, creative labor valuation, and algorithmic aversion, we theorize that AI involvement may trigger a perceived deficit in originality or intentionality, lowering perceived value. However, we also propose that transparency about the creative process can restore some of this lost value by reintroducing signals of effort and uniqueness. Using a VR-based mixed within-between subjects experiment, participants viewed artworks with systematically varied attribution and process information before expressing preferences. Our findings show a consistent “AI discount” in perceived value, but also reveal that this discount is significantly reduced — sometimes eliminated — when the AI-assisted creation is accompanied by high-effort disclosures or uniqueness cues. These results advance our understanding of how consumers make sense of creativity in the age of AI, and suggest that attribution and process transparency can meaningfully shape public reception of algorithmically generated content.

4 - Sustainable Procurement Contracts for Infrastructure Networks

Yuri Fonseca, Stanford University, Stanford, CA, United States, Vahideh Manshadi, Daniela Saban

In this work, we focus on the design and analysis of procurement mechanisms in which a set of products must be procured using one auction per product, yet these products have positive/negative synergies. Auctioneers often approach procurement as independent, especially when there is no overlap of bidders. For instance, possibly due to difficulty in coordination efforts, when developing infrastructure projects, the impact of each project for both society and the environment is often analyzed in isolation; however, when executed, synergies may occur, making the combined benefit super- (or sub-)additive of the individual ones. To understand the impact of ignoring the network effects on procurement, we characterize the optimal mechanism for the settings in which auctions are run simultaneously, or sequentially. In the simultaneous setting, each auction's optimal reserve price solves a pair of fixed-point equations reflecting how one project's success probability alters the marginal value of the other; positive externalities relax reserves, while negative ones tighten them. In the sequential setting, we show how reserves can be dynamically updated based on earlier outcomes. Next, we conduct a numerical study based on real infrastructure projects under Brazil's National Logistics Plan (PNL), demonstrating how ignoring these network effects in the design of the procurement auction leads to substantial loss across economic and environmental metrics, underscoring the importance of incorporating network effects into procurement design.

WA10

Building A Level 3 A312

Platform Design, Governance, and Ecosystem Dynamics

Contributed Session

Contributed

Chair: Luca Binder, Vlerick Business School, Gent, N/A, Belgium

1 - Beyond Zero-Sum Game: Deliberate Competition Building in Fragmented Markets

Abdullahi Babatunde, Africa Business School, Mohammed VI Polytechnic University, Rabat, Morocco, Dirk Boehe

This inductive, multiple-case qualitative study investigates why businesses in fragmented informal markets deliberately foster, rather than suppress, competition. Focusing on Nigeria's Igbo Apprenticeship System (IAS), where established entrepreneurs train apprentices to become future competitors, we introduce and theorize stewarded-competitions, defined as a culturally legitimized and economically motivated practice through which business owners deliberately develop new competitors by mentoring, capitalizing, and embedding them into shared market spaces. This distinct market logic challenges mainstream views on competition, especially the assumption that businesses succeed by protecting their territory or keeping out potential rivals. Based on 40 interviews across four industry sectors, we identify two connected drivers underpinning stewarded-competition: economic drivers, including keeping the market active and stable, sharing business risk across firms, and promoting long-term stability of the trading environment; and social-cultural drivers rooted in norms of reciprocity, communal responsibility, trust-based governance, and intergenerational continuity. Unlike cooperation frameworks, which depict collaboration as limited strategic alliances, stewarded-competition emerges as a culturally mandated yet strategically rational practice that proactively integrates economic sustainability and socio-cultural obligations. Our study contributes to neo-institutional theory by showing how informal institutions such as IAS go beyond filling institutional voids. They actively reconfigure competitive practices by embedding competition within culturally legitimate norms of mentorship, reciprocity, and collective responsibility. By theorizing stewarded-competition as a culturally grounded form of business logic, we extend the understanding of how informal institutions not only show how markets are organized but also how businesses compete and grow together in community-anchored informal economies.

2 - Strategic Focus in Digital Platforms: Capability Alignment for Sustainable Success

Michelle Vanassche, Vlerick Business School, Ghent, Belgium, Amit Basu, Steve Muylle

Digital platform capabilities are often studied in isolation from a platform's broader strategic focus. At the same time, many platforms are expanding their core functionality in an effort to become "super apps"; while this strategy offers opportunities for revenue diversification, it can backfire, as with X (formerly Twitter)'s pivot toward an "everything app". We present a conceptual model that links platform processes and capabilities to three distinct strategic activity types: transaction, engagement, and innovation. In contrast to prior research, our model emphasizes the importance of strategic focus, showing that platforms aligning their capabilities with a clear strategic direction are more likely to achieve sustainable success, whereas overly broad strategies tend to undermine platform performance. The model is validated by an extensive literature review and multiple case study research.

3 - Making It Epic? Demand-Side Reactions to Complementor-Driven Disruption in Platform Ecosystems

Luca Binder, Vlerick Business School, Ghent, Belgium, Ksenia Podoynitsyna

Digital platform ecosystems are increasingly shaped not only by platform owners but by strategic moves of powerful complementors. However, little is known about how end-users respond to such disruptions - particularly in mature platform markets where network effects and high switching costs are assumed to insulate incumbents. This study explores demand-side reactions to ecosystem fragmentation by analyzing end-user discourse surrounding Epic Games' adjacent entry with the Epic Games Store (EGS) into PC game distribution, a market historically dominated by Valve's Steam platform.

Drawing on high-engagement Reddit threads from 2017 to 2024, we employ transformer-based topic modeling (BERTopic) to trace discursive shifts and sentiment patterns across four distinct phases of Epic's ecosystem disruption. Our analysis captures how perceptions of both actors evolved in response to value bundling (e.g., cross-platform features) and value fragmentation (e.g., exclusive content), revealing iterative patterns in legitimacy assessments and multi-homing intentions.

Building on these insights, we develop a process model that traces how complementor disruption sets in motion a dynamic interplay between complementor platformization, incumbent repositioning, and end-user coordination. The model highlights pathways through which disruptive entrants and incumbent platforms jointly reconfigure ecosystem boundaries and how powerful complementors can position themselves as alternative platform gatekeeper.

Theoretically, we contribute to research on platform ecosystem evolution by proposing a process model of complementor disruption that foregrounds end-user mobilization as a key mechanism in shaping the trajectory of fragmented platform markets. Methodologically, we illustrate the value of machine learning-based text analysis for inductive theorizing from large-scale user discussions.

WA11

Building A Level 3 A313

Experiment Design and Causal Inference in Operations

Invited Session

Revenue Management

Chair: Sentao Miao, University of Colorado Boulder, Westminster, United States

Co-Chair: Zhiyuan Tang, The University of Texas at Dallas, Richardson, TX, United States

1 - Experimental Design in Live-Interaction Platforms

Chenran Weng, University of California, Berkeley, Berkeley, CA, United States, Xiao Lei, Nian Si

We investigate experimental design problems in live-interaction platforms, which are commonly found in the online gaming industry and anonymous social networks. These platforms involve users being matched with others for activities such as games or social interactions, resulting in an interdependence of users' metrics on the treatment assignments of their counterparts. We construct a stochastic market model and develop its mean field limit to analyze the experimental dynamics in such platforms. Our focus is on two randomization strategies: user randomization and match randomization. We demonstrate that under Markovian user behavior and

homogeneous treatment effects, match randomization can provide unbiased estimations. However, significant biases may arise when these conditions are not met. On the other hand, user randomization is generally biased but exhibits greater resilience to model inaccuracies. We then propose a simple linear regression estimator under user randomization and demonstrate that this estimator consistently outperforms alternatives in various situations.

2 - Disentangling Interference of Multiple Promotions: a Debiased Machine Learning Approach

zhe wang, Boston University, Allston, MA, United States, Jinglong Zhao

Online platforms send multiple promotions each day, which can lead to complex interference that makes it difficult to understand the impact of each individual promotion. In this paper, we adopt a debiased machine learning approach to study the interference problem arising from the multiple promotions.

3 - Debiased Difference-in-Differences Estimation and Inference of Causal Effects Under Adaptive Controls

Zhiyuan Tang, The University of Texas at Dallas, Richardson, TX, United States, Sentao Miao, Yining Wang

As adaptive experiments and data-driven decision systems become increasingly prevalent in managerial contexts, new challenges emerge regarding the validity of difference-in-differences (DiD) methods when the data collection process is influenced by prior outcomes. For example, panel data may be gathered during a dynamic pricing experiment in which prices are set adaptively based on historical prices and customer responses. In such settings, key covariates—such as price—is governed by an underlying policy that induces dependence across observations. This violates standard DiD assumptions and introduces bias into conventional estimators.

In this paper, we address the problem of estimating and conducting valid inference on causal effects within the DiD framework when data are adaptively collected. Our contributions are three-fold. First, we develop a general debiasing method that converts the bias of standard DiD estimators into variance, enabling valid inference under adaptive data collection. Second, we derive finite-sample bounds on the bias and variance of the proposed estimator and establish the asymptotic validity of the associated confidence intervals. Third, we demonstrate the applicability of our method through a dynamic pricing example, showing that our estimator achieves a minimax-optimal trade-off between estimation accuracy and policy regret. Taken together, our results provide a robust framework for applying DiD method in sequential decision-making environments, extending its applicability to modern adaptive data settings.

4 - Policy-Aware Design of Large-Scale Factorial Experiments

Xin Wen, New York University, New York, NY, United States, Xi Chen, Will Wei Sun, Yichen Zhang

Modern digital platforms and policy evaluations increasingly face “graduated A/B tests,” where dozens of features or interventions—each with multiple dosage levels—must be assessed jointly. Exhaustively testing every combination is prohibitively costly, while ad-hoc pruning risks discarding the best option. We study policy-aware design for such large-scale factorial experiments. Building on the observation that treatment–response tensors are typically low-rank, we develop a two-stage framework that (i) learns a compact latent-factor representation of combinatorial effects from a pilot sample and (ii) adaptively allocates additional units to identify a near-optimal treatment bundle. The resulting algorithm balances exploration and exploitation without presuming additive effects or requiring strong ignorability across concurrent experiments. Simulation evidence shows that, for realistic traffic budgets, our approach attains the same error guarantees as full-factorial designs with an order-of-magnitude fewer observations. An illustrative platform application highlights how the method can shorten experimentation cycles while remaining transparent to decision-makers. Beyond digital products, the framework is applicable to multi-component policy packages and clinical treatment regimens, offering a practical path toward evidence-based implementation when the search space is combinatorially large and ethical or opportunity costs constrain sample size.

WA12

Building A Level 3 A314

Policy-Enabling Models in the Electricity Sector

Invited Session

ENRE: Electricity

Chair: Afzal Siddiqui, Stockholm University, Kista, Sweden

Co-Chair: Ramteen Sioshansi, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Uniqueness of Nash-Cournot Equilibria in Wholesale Electricity Markets with Variable Renewable Energy and Demand Uncertainty

Joseph Duggan, University of Dayton, Dayton, OH, United States

This work explores the uniqueness of Nash-Cournot equilibria in wholesale electricity markets with both thermal and renewable generation in the presence of both uncertain demand and variable renewable generation. In a previous work (Chaiken and Duggan (2024)), the authors expanded upon the uniqueness results proven by Lagerl (2006) to incorporate the presence of renewable generation into the market in the manner of Acemoglu et al. (2017). The model of Acemoglu et al. (2017) incorporates a constant amount of renewable generation, with zero marginal cost, distributed among all firms. The present work extends this model to include uncertainty over the amount of renewable generation available in any given period, and extends the prior uniqueness results to study the effects of variable renewable supply on the multiplicity of equilibria.

2 - Modeling Day Ahead Ancillary Services Initiative (Dasi)

Golbon Zakeri, University of Massachusetts - Amherst, Amherst, MA, United States

Energy imbalance reserve (EIR) product is introduced into the Independent System Operator (ISO) of New England's day-ahead wholesale electricity market to provide a better fuel procurement incentive for generating resources. Different from existing forward reserve products, EIR is a novel real option product, which is settled against real-time energy price rather than reserve prices. In this presentation, we develop a stochastic equilibrium model that incorporates the risk preference of generator and demand agents participating in the energy and reserve market in both day-ahead and real-time time frame.

3 - Sustainable and Reliable Pathways to Electricity Capacity Auctions

Emily Cai, Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, United States, Alan Scheller-Wolf, Ramteen Sioshansi

This ongoing project studies the joint behavior of regulators, renewable generators, and non-renewable generators in modern electricity markets to identify pathways for securing sustainable and reliable generation resources. Procuring adequate capacity for future electricity demand is a challenging and critical task for independent system operators and regional transmission regulators. Doing so while incentivizing the right balance between green and reliable firm power generation resources is nontrivial. How can regulators secure reliable and cheap capacity, with increasing electricity demand on one hand and public policies incentivizing renewable energy generation on the other? Equally important, how do renewable and non-renewables react to these rules? Why, when and how should renewables participate in capacity auctions? We consider an oligopolistic setting with one representative renewable generation firm, one non-renewable generation firm, a load-serving entity (utility company), a regulator, and a constrained locational deliverability area. These market participants interact with one another in three connecting markets—capacity, spot, and renewable energy credit. We intend to address three fundamental issues: 1) Modeling the capacity planning problem for renewable power generation resources based on the interaction between auction mechanisms and public policies, 2) Aligning the foundational objective of a capacity auction to secure reliable generation capacity with environmental health and sustainability efforts, and 3) Developing operational and business insights that guide power generation stakeholders—consumers, load serving entities, and generation resources—towards a greener, more sustainable, and reliable energy future.

4 - Integrating Floating Offshore Wind In Capacity Expansion Planning: a California Case Study

Jessica Wert, Lawrence Livermore National Laboratory, Livermore, CA, United States, Jean-Paul Watson, Matthew Signorotti, Robert Arthur, Hsiang-He Lee, Jean-Christophe Golaz

Diversifying energy sources supports grid resilience, and California's proximity to strong, relatively uniform winds off the coast presents floating offshore wind as a promising option. This study integrates floating offshore wind into Lawrence Livermore National Laboratory's (LLNL) nodal capacity expansion planning model to evaluate bounded and cost-effective buildouts of generation, storage, and transmission infrastructure under future conditions. Using three cost trajectories and 100 representative days informed by high-resolution climate data from DOE's E3SM climate model, the stochastic mixed-integer optimization model assesses the feasibility and cost-effectiveness of offshore wind in California's grid.

WA13

Building A Level 3 A315

Machine Learning in Optimization

Contributed Session

Contributed

Chair: Yue Lin, The Chinese University of Hong Kong, Hong Kong, Hong Kong

1 - Federated Learning for Stochastic Scenario Generation with Cost-Aware Prediction

Mehrdad Mohammadi, Auburn University, Auburn, AL, United States, Alexander Vinel

This study presents a federated learning framework for generating stochastic scenarios in a decentralized and privacy-preserving manner. The approach captures data heterogeneity without raw data sharing by training models locally across clients and aggregating them globally. The model incorporates prediction accuracy (via MAE and MAPE) and operational cost metrics (holding, backorder, and ordering) to ensure realistic and decision-relevant scenarios. This cost-aware, federated optimization method enhances decision-making under uncertainty and demonstrates strong potential for applications in data-sensitive environments.

2 - Stochastic Qubo Optimization via Bayesian Factorization Machines and Quantum Annealing

Tomah Sogabe, The University of Electro-Communications, Chofu, Japan, Yuya Ito, Taihei Kuroiwa, Daiki Yamazaki, Kiria Yamashita, Daiki Sasaki, Ryosuke Koga

Combinatorial optimization under uncertainty remains a challenging problem across a wide range of domains, including logistics, finance, and energy systems. Traditional deterministic approaches and scenario-based stochastic programming methods often fail to capture deep correlations among uncertain parameters, leading to solutions that are either costly or risk-prone. In this study, we propose a probabilistic optimization framework called BFMQA, which integrates Bayesian Factorization Machines (BFM) with Quantum Annealing (QA). BFM represents costs, demands, and risk factors as a sparse interaction matrix and learns complex probabilistic correlations by estimating the posterior distribution of latent factors. The resulting probabilistic model is directly embedded into a Quadratic Unconstrained Binary Optimization (QUBO) formulation, enabling the integrated expression of economic objectives and risk penalties. QA then explores the high-dimensional optimization landscape in a quantum superposition state, efficiently generating candidate solutions that account for correlated uncertainties while avoiding local optima. Numerical experiments demonstrate that the proposed method achieves high consistency with solutions obtained by conventional deterministic and scenario-based stochastic programming approaches, while significantly reducing computation time and avoiding excessive risk. Thanks to its modular design, this framework can be easily extended to a variety of domains such as supply chain design and risk-aware portfolio management, offering a promising path toward the practical realization of uncertainty-resilient optimization using quantum hardware.

3 - Robust Data-Driven Cara Optimization

Li Chen, The University of Sydney, CHIPPENDALE, Australia, Arjun Ramachandra, Napat Rujeerapaiboon

Problem definition: Inspired by Von Neumann's foundational work on expected utility, this paper extends the data-driven expected utility maximization framework under constant absolute risk aversion (CARA) preferences by incorporating nonlinear payoff functions and enhancing robustness. These improvements aim to broaden the applicability of utility maximization in prescriptive analytics and to improve out-of-sample performance, particularly in data-scarce settings, respectively.

Methodology/results: The combined nonlinearity of exponential utility and payoff functions presents significant computational challenges in ensuring robustness. We show that existing safe tractable approximation techniques are overly conservative, potentially causing unexpected infeasibility. To address this, we define a notion of consistent approximation and propose an augmentation technique that provably retains the consistency in the approximation.

Managerial implications: Theoretical insights are validated through computational experiments in two applications: data-driven portfolio optimization and facility location optimization. Our results demonstrate the effectiveness of incorporating robustness into prescriptive analytics, yielding solutions that significantly outperform traditional empirical and stochastic optimization. Our utility maximization framework is therefore not only robust, but also interpretable, scalable, and adaptable to a wide variety of real-world risk-aware decision-making problems involving nonlinear payoffs.

4 - Faster Global Minimum Cut with Predictions

Helia Niaparast, Carnegie Mellon University, Pittsburgh, PA, United States, Benjamin Moseley, Karan Singh

Global minimum cut is a fundamental combinatorial optimization problem with wide-ranging applications. Often in practice, these problems are solved repeatedly on families of similar or related instances. However, the de facto algorithmic approach is to solve each instance of the problem from scratch discarding information from prior instances.

In this paper, we consider how predictions informed by prior instances can be used to warm-start practical minimum cut algorithms. The paper considers the widely used Karger's algorithm and its counterpart, the Karger-Stein algorithm. Given good predictions, we show these algorithms become near-linear time and have robust performance to erroneous predictions. Both of these algorithms are randomized edge-contraction algorithms. Our natural idea is to probabilistically prioritize the contraction of edges that are unlikely to be in the minimum cut.

5 - Asymptotically Optimal Distributionally Robust Solutions Through Forecasting and Operations Decentralization

Yue Lin, Chinese University of Hong Kong, Hong Kong, Hong Kong, Daniel Zhuoyu Long, Viet Anh Nguyen, Jin Qi

Two-stage risk-averse distributionally robust optimization (DRO) problems are ubiquitous across many engineering and business applications. In these problems, decision-makers commit to capacitated first-stage decisions, anticipating that they can execute second-stage recourse decisions after observing the realization of the uncertain parameters. Despite their promising resilience, two-stage DRO problems are generally computationally intractable. To address this challenge, we propose a simple framework by decentralizing the decision-making process into two specialized teams: forecasting and operations. This decentralization aligns with prevalent organizational practices, in which the operations team uses the information communicated from the forecasting team as input to make decisions. We formalize this decentralized procedure as a bilevel problem to design a communicated distribution that can yield asymptotically optimal solutions to the original two-stage risk-averse DRO problems. We identify an optimal solution that is surprisingly simple: The forecasting team only needs to communicate a two-point distribution to the operations team. Consequently, the operations team can solve a highly tractable and scalable optimization problem to identify asymptotically optimal solutions. Specifically, as the magnitude of the problem parameters (including the uncertain parameters and the first-stage capacity) increases to infinity at an appropriate rate, the cost ratio between our induced solution and the original optimal solution converges to one, indicating that our decentralized approach yields high-quality solutions. We compare our decentralized approach against the truncated linear decision rule approximation, illustrating its broader applicability and superior computational efficiency with competitive performance. The finely tuned solution outperforms traditional sample-average approximation methods in out-of-sample performance, underscoring its practical viability.

WA14

Building A Level 3 A316

Exact and Heuristic Methods for Advanced Vehicle Routing

Contributed Session

Contributed

Chair: Xinbo Zhang, City University of Hong Kong, Room 5216, OBOR Center, Lau Ming Wai Building., Hong Kong, Hong Kong

1 - The Drone Scheduling Problem in Shore-to-Ship Delivery: A Time Discretization-Based Model with an Exact Solving Approach

Xiaodeng Hao, Hong Kong Polytechnic University, Hong Kong, Hong Kong, Ying Yang, Silong Zhang, Zhou Xu

Amid growing interest in the integration of drones into maritime logistics, this paper addresses the drone scheduling problem in shore-to-ship delivery (DSP-SSD). We introduce a time-discretized mixed-integer programming model incorporating drone-specific constraints and moving targets. For larger instances, we propose a tailored branch-and-price-and-cut (BPC) algorithm, which integrates a specialized backward labeling algorithm, cutting planes, and acceleration methods. Experimental results demonstrate that our BPC algorithm substantially outperforms commercial solvers in both solution quality and computational efficiency. We also conduct sensitivity analyses on key parameters to provide practical insights for real-world DSP-SSD operations.

2 - An Exact Price-and-Enumerate-and-Cut Algorithm for Vehicle Routing Problem with Speed Optimization

Ying Yang, Hong Kong Polytechnic University, Hong Kong, Hong Kong, Silong ZHANG, Shuaian Wang

We study the vehicle routing problem with speed optimization (VRP-SO), in which vehicle routes and travel speeds are jointly optimized to minimize total operational costs. These costs are modeled as a convex function of speed, capturing key trade-offs such as fuel consumption and driver wages. The simultaneous consideration of routing and continuous speed decisions introduces significant computational challenges.

To solve the VRP-SO, we develop an exact price-and-enumerate-and-cut (PEC) algorithm. The PEC algorithm relies on a tailored set-partitioning formulation in which each column represents a complete route with an associated speed profile. We present a novel labeling algorithm for solving the pricing subproblem, which employ closed-form analytical expressions for optimal speeds and efficient dominance rule. These eliminate the need to solve convex optimization problems during label extension and significantly improves computational performance while maintaining exactness. Additionally, we introduce a dedicated route enumeration phase that systematically generates a pool of high-quality, speed-optimized routes. This framework eliminates the need for branching on integer variables and is enhanced by valid inequalities and effective bound-tightening techniques.

We evaluate the PEC algorithm on both pollution routing and maritime transportation benchmark instances. Computational results demonstrate that the algorithm efficiently solves small- and medium-scale instances and consistently outperforms existing methods in terms of both runtime and solution quality.

3 - A Two-Stage Exact Method for the Fair Capacitated Vehicle Routing Problem

Dong Han, University of Florida, Gainesville, FL, United States, Zhengzhong You, Yu Yang

Fairness in optimization has gained increasing attention, with both cost efficiency and equitable outcomes recognized as critical objectives. A common fairness measure is the range, defined as the difference between the largest and smallest values of interest. In this study, we focus on the Fair Capacitated Vehicle Routing Problem (F-CVRP), which aims to find a feasible set of routes that minimizes the range of route distances. To efficiently handle the computational challenges of F-CVRP, we propose a two-stage solution framework. In the first stage, we apply modern techniques designed for CVRP and use RouteOpt, a state-of-the-art solver, to shrink the feasible region of the original problem, keeping only a small number of routes that can be explicitly enumerated while ensuring optimal solution to F-CVRP is kept. In the second stage, we solve a mixed-integer programming (MIP) model using these enumerated routes to find the optimal solution. To improve tractability, we introduce a set of valid inequalities and develop a tailored branching strategy. Our computational results show that the proposed approach can solve F-CVRP instances with up to 60 customers, which is three times larger than the instance sizes solved in the most recent literature, to the best of our knowledge.

4 - An ALNS-Based Framework for the Capacitated Vehicle Routing Problem with A Zone Tariff

Manuel Ostermeier, University of Augsburg, Augsburg, Germany, Niklas Tuma, Alexander Hübner

Cost pressure and fierce competition force companies to focus on their core competencies. Therefore, 90% of Fortune 500 companies in the United States outsource parts of their distribution planning to logistics service providers (LSPs). LSPs help to buffer peak demands and offer cost savings through economies of scale. This trend also applies to the retail industry. Most retailers supply their stores with less-than-truckload shipments from a distribution center. They propose store clusters to the LSPs, which check and execute the tours. The LSPs bill according to a zone tariff where the tour price depends on the farthest zone visited and the carried load, including volume discounts. Zone tariffs reduce the planning complexity for retailers since the clustering and pricing are independent of the delivery sequence.

To minimize costs, retailers must form clusters that respect capacity and detour limits, which are specific distance constraints. We formalize the resulting planning problem as a Capacitated Vehicle Routing Problem with a Zone Tariff (C-VRP-ZT). Zone tariffs are widely applied in practice but mostly neglected in research. Existing C-VRP-ZT approaches focus on exact approaches for small instances and an absolute detour limit. We propose an ALNS-based matheuristic framework using problem-specific operators and multiple adaptive mechanisms to solve realistic (large) problem instances and different detour limitations. The algorithm comprises a heuristic tour generation component and an exact post-optimization component. We provide extensive numerical studies on datasets up to 1000 nodes, including real-world data from an industry partner.

5 - An Exact Algorithm for the Traveling Salesman Problem with Pickup and Delivery Under Time Windows and Ride-Time Constraints

Xinbo Zhang, City University of Hong Kong, Hong Kong, Hong Kong, Roberto Baldacci

Traveling salesman problem with pickup, delivery, and ride-time constraints (TSPPD-RT) requires a depot-based vehicle to visit multiple known requests. Each request specifies a pickup vertex, corresponding delivery vertex, and maximum travel time between them (ride-time). The goal is to find a tour that (i) starts and finishes at the depot, (ii) visits all requests, (iii) satisfies ride-time constraints for each request, and (iv) minimizes the vehicle's total travel cost. Additionally, undelivered requests must remain within the vehicle's capacity limit. The TSPPD-RT is a fundamental optimization problem in logistics and routing. This paper introduces an exact method for solving an extended version of the TSPPD-RT with time window constraints (TSPPD-TWRT), incorporating both soft and hard ride-time constraints (RDCs). RDCs consider only travel times in the soft case, while in the hard case, they also include waiting times. This paper presents two mathematical formulations: an arc-based and path-based, with the latter derived from the U-tour relaxation. Using these formulations, we develop MIP-based and branch-price-and-cut solution approaches. Computational experiments on datasets from the Dial-A-Ride Problem and the Pickup and Delivery Problem with Time Windows demonstrate the effectiveness of the proposed methods. Results indicate that the U-tour formulation is particularly efficient, solving instances with up to 60 requests within reasonable computational times.

WA15

Building A Level 4 A401

Demand Modeling in Urban Network

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Youngseo Kim, UCLA, Santa Monica, CA, United States

1 - Ride-Sourcing Demand Estimation Using Aggregated Public Data

Yuxin Sun, University of Massachusetts, Amherst, Amherst, MA, United States, Song Gao

The study of the ride-sourcing system is inevitably constrained by the availability of data owned by private platforms. With a few exceptions, public ride-sourcing data is highly aggregated. In this paper, we develop a methodology to estimate ride-sourcing demand at a high spatial resolution (e.g., census tract or higher) based on low-resolution public data, such as the town-level dataset from Massachusetts, making use of geographically weighted regression (GWR) that has been shown to produce robust estimation results across various spatial aggregation levels. The approach uses publicly available socioeconomic, demographic, land use, and built environment data to estimate two sets of models: (1) GWR trip origin (production) and destination (attraction) models that account for the spatial heterogeneity in underlying relationships; and (2) a Multinomial Logit model for the distribution of trips to different destination zones for a given origin. Our analysis reveals significant spatial dependence in ride-sourcing trips, highlighting the influence of local travel behavior patterns, such as commuting and job distribution. By incorporating Origin-Destination Employment data, our framework captures both commute and non-commute factors, offering a more comprehensive understanding of various ride-sourcing trip purposes. Our results provide a scalable framework for exploratory and predictive analyses of ride-sourcing demand patterns. The estimated census-tract level demand data will be put in a public data repository for the research community and beyond.

2 - Reduced-Basis Method for Fast Solutions of Traffic Flow Problems

Ali Pirmohammadi, New York University, Brooklyn, NY, United States, Daniel Vignon

A number of situations require repeated and/or real-time solutions of traffic flow equations in a network. Such scenarios include network design problems in which the underlying network is dynamic and the coordination of traveler movement and emergency response teams during a network-wide disruption. In the former case, the choice of an appropriate design requires the computation of link flows for many designs. In the latter case, one requires fast decisions from system managers while historical data is not applicable. In many systems, model order reduction methods exist to reduce the computational burden, but their application to traffic network problems is limited due to the hyperbolic nature of the partial differential equations (PDEs) and the high-dimensional nature of the many inputs (initial, and boundary conditions). In this study, to overcome the inherent reduction challenges posed by hyperbolic PDEs, we apply the reduced-basis method (RB) to reduce the viscosity solution of a parameterized first-order traffic flow PDE. This approach enables us to construct a solution manifold approximated by a small set of linear basis functions, facilitating rapid solutions for specified time horizons, boundary, and initial conditions. Both the offline training stage and the online stage are made computationally efficient by using an effective greedy sampling method and the Discrete Empirical Interpolation Method (DEIM). To allow for fast solutions at the network level, we extend our framework to handle the reduction of coupled PDEs that describe interactions at junction. Finally, we demonstrate how our methodology performs in the context of network design problems.

3 - Modeling Realized Accessibility in Space&TIME Networks: Why Do We Avoid Places We Can Reach?

Hiroyuki Hasada, The University of Tokyo, Meguro-ku, Japan, Yudai Honma, Xuesong Zhou

Realized accessibility, defined as the extent to which individuals actually visit locations, often diverges from what is theoretically accessible within spatial and temporal constraints. For instance, a person may travel 3 kilometers to visit a large shopping mall with diverse stores and amenities, but choose not to visit a smaller commercial street at the same distance. This study aims to explain the persistent gap between theoretical and realized accessibility using a novel modeling framework on a space-time network, where each link corresponds to either travel or stay segments. We formulate an inverse shortest path problem (ISPP), which recovers link costs that make observed travel paths appear optimal. The formulation allows us to infer underlying preferences without explicitly enumerating the choice set, which is especially valuable on space-time networks with many possible paths. When the model is applied to simple out and back trips, its structure naturally aligns with the formulation of a soft margin support vector machine (SVM), offering interpretable decision boundaries in accessibility space. In real-world urban behavior, our approach captures how individuals choose to stop at specific facilities during their daily tours, revealing patterns that depend on the temporal availability and the perceived attractiveness corresponding to each stay link cost. The optimized link costs vary systematically with facility type, time of day, and individual characteristics. Even when theoretical accessibility is identical, realized accessibility differs significantly between individuals or residential locations. Tensor-based computation accelerates estimation and improves clarity, enabling scalable learning from multidimensional behavioral data.

4 - Integrating Qualitative Data into Decision-Making for Transportation and Logistics

Shriya Karam, Massachusetts Institute of Technology, Cambridge, MA, United States, Alexandre Jacquillat

Public transit agencies face challenges in aligning supply levels with evolving passenger demand. To better understand passenger demand patterns and improve service design, some agencies have begun exploring qualitative data — such as social media posts or customer comment cards — to provide insights on passenger experiences. Yet, unstructured feedback is rarely incorporated into tactical or strategic decision-making, leaving a potentially valuable source of demand information underused. To address this gap, we develop a methodology that leverages qualitative data to derive operational value, combining natural language processing, data-driven demand estimation, and stochastic optimization. As a case study, we use Tweets and customer comment cards from the Washington Metropolitan Area Transit Authority to re-optimize rail timetables. Our two-step approach first combines qualitative customer feedback with aggregated numerical demand data to estimate origin-destination demand matrices. In the second step, we use these demand estimates to re-optimize transit frequencies via a two-stage stochastic optimization problem. Ultimately, our results demonstrate that our demand estimation approach outperforms benchmark models and highlight the practical impact of incorporating qualitative data into transportation decision-making.

5 - Estimate Then Predict: Convex Formulation for Travel Demand Forecasting

Youngseo Kim, UCLA, Santa Monica, CA, United States, Gioele Zardini, Samitha Samaranyake

Travel demand forecasting models play crucial roles in evaluating large-scale infrastructure projects, such as the construction of new roads or transit lines. While combined modeling approaches have been explored as a solution to overcome the problem of input/output discrepancies in a sequential four-step modeling process, previous attempts at combined models have encountered challenges in real-world applications, primarily due to their limited behavioral richness or computational tractability. In this study, we propose a novel convex programming approach and present a key theorem demonstrating that the produced optimal solution coincides with the one arising from the hierarchical extended logit model. This model is specifically designed to capture correlations existing in travelers' choices, including similarities among transport modes and route overlaps. The convex property of our model ensures the existence of solutions and naturally offers computational efficiency. The advantages of our proposed model are twofold. First, it provides a single unifying rationale (i.e., utility/entropy maximization) that is valid across all steps. Second, its combined nature allows one to systematically handle observed data, enabling a better representation of reality. Our convex programming approach shows promise in enhancing the accuracy and applicability of travel demand forecasting, thereby aiding in the decision-making processes for infrastructure projects.

WA16

Building A Level 4 A402

AI and Data-Driven Strategies in the Future of Air Mobility

Invited Session

Air Transportation Section

Chair: Soohwan Oh, N/A

1 - Active Risk Mitigation for Unmanned Aerial Systems

Andrew Kendall, University of Texas at Austin, Austin, TX, United States

As Unmanned Aerial Systems (UAS) enter routine cargo operations within the National Airspace System (NAS), quantitative approaches are needed to assess and mitigate operational risk. Because aircraft accidents are rare but high-consequence events, conventional Monte Carlo simulation is often inefficient for estimating their probabilities.

This talk presents a simulation-based optimization framework for active risk mitigation in UAS operations. An agent-based model captures interactions among aircraft systems, environmental disturbances, and decision-making agents during critical flight phases. To estimate accident probabilities efficiently, a Metropolis-Adjusted Langevin Algorithm (MALA) is applied for rare-event sampling, enabling exploration of high-dimensional failure scenarios involving navigation errors, wind shear, and engine failures.

The approach integrates rare-event estimation with adaptive decision policies that allocate computational effort toward high-risk regions of the state space, improving statistical efficiency while supporting real-time safety management. Results demonstrate how the method enables risk-informed decision-making consistent with FAA safety objectives and provides a generalizable framework for agent based autonomous decision-making in safety-critical domains, including air traffic management, logistics, and infrastructure reliability.

2 - He B-737 Max Saga: Cost Implications of Safety Management in Aircraft Production

Dr I Richmond Nettey, College of Aeronautics and Engineering, Kent State University, Kent, OH, United States

The loss of 346 crew and passengers onboard Lion Air Flight 610 (JT610) on 29th October 2018, and Ethiopian Airlines Flight 302 (ET302) on 10th March 2019, resulted in grounding of the Boeing 737 MAX aircraft between March 2019 and December 2020. These two Boeing 737 MAX 8 accidents, “have cost Boeing more than \$20 billion in compensation, production costs, and fines, led to a 20-month grounding for the bestselling plane (Shepardson, [Reuters](#), 28th November 2023).

Alaska Airlines Flight 1282, a Boeing 737 MAX 9, suffered cabin panel blowout after departure from Portland International Airport on 5th January 2024, while climbing through 14,800’ MSL. If the blowout had occurred at a higher altitude, the losses would have been tragic. On 25th March 2024, Boeing’s leadership team of the CEO and head of airplane production, announced their departure.

These tragic losses underscore safety’s importance in the aviation, manufacturing, and service sectors of the economy. “As with every major accident, the lessons learned go far beyond addressing the primary causal factors. With two accidents to learn from, . . . , the MAX story’s influence on aircraft design, certification, and pilot training is only starting to be felt” (Broderick, 2020). This presentation will examine the importance of safety data through the deployment, design, manufacturing, and production of the Boeing 737 MAX. Specifically, it will examine: development of the B-737; competition for market share; safety challenges in the B-737 MAX development; and costs and implications for Boeing and the airlines.

3 - Development of Decision Support Systems for Advanced Air Mobility Operations in South Korea

Junghyun Kim, Sejong University, Seoul, Korea, Republic of

The Korean government has started exploring the implementation of Advanced Air Mobility (AAM) due to its potential to enhance connectivity in both urban and regional areas. The implementation of AAM requires careful decisions to address critical challenges such as vertiport placement and corridor network design. This paper presents the development of decision support systems for policymakers to evaluate various operational strategies for the upcoming AAM operations in South Korea. Three interconnected systems are developed by leveraging a data-driven approach: unsupervised learning-based vertiport placement, dynamic corridor network generation, and flight path planning. Each system uses relevant datasets tailored to its specific purpose and employs algorithms to enable the proposed concepts. The main contributions of this research are to 1) propose an unsupervised learning-based approach that can be utilized for identifying optimal vertiport locations in the Korean peninsula, 2) establish a dynamic corridor network generation system, and 3) develop a multi-objective flight path optimization system customized to AAM operations in South Korea. The authors believe that these systems can be used to assess various operational initiatives, particularly for single flight scenarios; therefore, policymakers can make informed decisions for effective implementations of the upcoming AAM operations in South Korea.

4 - Data-Driven Spatiotemporal Risk Modeling for Advanced Air Mobility

Soohwan Oh, Hanseo University, Chungcheongnam-do, Korea, Republic of

This study presents a data-driven framework for modeling spatiotemporal risks for advanced air mobility operations. By replacing static census data with high-resolution hourly de facto population data for Seoul, we compute a risk index that transforms dynamic restricted airspace zones into acceptable safety levels. The daytime maps show a dense concentration of restricted areas over commercial areas, while residential areas are largely visible, a pattern that is often overlooked when using traditional residential data.

WA17

Building A Level 4 A403

Future of Work: AI Automation, and Human Collaboration

Invited Session

Practice Track

Chair: Beverly Wright, Wavicle Data Solutions, Anderson, SC, United States

1 - Utilizing AI to Enhance the Claims Audit Process

Samuel Damon, Riverstone, Manchester, NH, United States, Temi Gaudet

The claims audit process is integral to ensuring compliance, operational efficiency, and customer satisfaction within the insurance industry. However, traditional methods often struggle to scale effectively due to the complexity and volume of claims data. This presentation introduces an innovative approach using agents built on large language models (LLMs) to transform the claims audit workflow. These AI-powered agents leverage pre-existing claims audit guides to break down audit questions, determine their query feasibility, locate relevant data sources, analyze results, and assess whether claims were handled satisfactorily. By automating and optimizing key steps in the audit process, this solution significantly reduces manual effort while increasing accuracy and consistency. Focus will highlight the technical implementation, challenges addressed (such as data integration and context understanding), measurable outcomes, including faster audits and improved decision-making as well as multiple business value streams derived from the effort.

2 - The Agility Revolution: Harnessing AI Agents to Turn Market Challenges into Competitive Advantage

Asad Musa, Innova Solutions, Alpharetta, GA, United States

In an era of relentless market volatility and hyper-accelerated competition, organizations face a stark ultimatum: evolve at the speed of disruption or fade into irrelevance. Enter the agility revolution, powered by AI agents—intelligent systems that don’t just react to change but anticipate and shape it. This panel unveils how trailblazing organizations are harnessing AI agents to transform market challenges into bold competitive advantages, redefining the art of strategy in a digital age. With this fast-evolving revolution, it raises provocative questions: How do we craft AI agents that amplify human ingenuity rather than supplant it? What becomes of corporate culture when autonomous systems drive split-second market responses beyond human control? Through vivid case studies from diverse industries, this presentation reveals how leading firms are building “predictive agility”—the capacity to foresee market shifts and act decisively before opportunities vanish.

3 - Discuss Reflections and Thoughts About the Future of Work

Lakshman Kaveti, ProArch Technologies, Suwanee, GA, United States

Discuss reflections and thoughts about the future of work

4 - Humanity in Flux: Navigating Identity, Ethics, and Value in the Age of AI

Beverly Wright, Wavicle Data Solutions, Anderson, SC, United States

As artificial intelligence increasingly influences the ways we live, work, and connect, it is reshaping not just our systems – but our sense of what it means to be human. This panel will explore how rapid technological advancements challenge our long-standing views of identity, creativity, trust, and ethics. How do individuals and organizations preserve humanity in the face of automation? What new skills, mindsets, and ethical frameworks are essential for thriving in this dynamic environment? Join this discussion about the shifting relationship between people and machines. We'll examine real-world examples from business, education, healthcare, and society – highlighting both the risks of dehumanization and the unprecedented opportunities for human flourishing. This conversation will spark ideas and strategies for leading with empathy, designing for trust, and fostering meaningful human connections in a world increasingly shaped by AI.

WA18

Building A Level 4 A404

Intelligent Infrastructure and Urban Systems

Contributed Session

Contributed

Chair: Amin Keramati, Widener University, New York, NY, United States

1 - Using MLP for Capturing Drivers' Preferences for Improving Adherence in a VRP Optimization Setting

Luis Aburto, Universidad Adolfo Ibáñez, Santiago, Chile, Felipe Lagos, Víctor Bucarey, Tomás Rubio

The Vehicle Routing Problem (VRP) is a well-known problem in the operations research literature, traditionally addressed by minimizing transportation costs. However, in practice, cost matrices are often outdated or fail to reflect real-world conditions. Additionally, drivers frequently develop personal routing preferences based on their daily experience, which are typically ignored in standard VRP models. This project aims to bridge that gap by leveraging GPS data to capture driver route preferences. These preferences are then used to train a multilayer perceptron (MLP) to predict preferred routes based on the specific characteristics of a VRP instance, particularly customer locations. The predicted preferences are integrated into a VRP optimization model to adjust route recommendations and customer-to-vehicle assignments accordingly. We apply this methodology in collaboration with a VRP service provider in Latin America. Empirical results show a significant improvement in route adherence, increasing from 33.6% to 52.9%. This work presents a novel integration of machine learning and optimization techniques to enhance solution quality and incorporate user preferences in logistics planning.

2 - Reinforcement Learning Based Framework to Design Behavioral Interventions to Improve Sustainability and Public Health

Viswa Sri Rupa Anne, Georgia Institute of Technology, Atlanta, GA, United States, Yuming Chang, Srinivas Peeta

Transportation systems in the U.S. are a major source of air pollutants and greenhouse gas emissions, resulting in a growing interest to promote sustainability, public health and well-being. Promoting shifts towards sustainable modes (i.e., walking, biking, or transit) or routes can reduce greenhouse gas emissions, improve air quality, and increase physical activity levels. To promote such shifts, behavioral interventions such as monetary rewards, informational messages communicating health benefits, and gamification strategies have emerged as potential solutions to influence travel behavior. Mobile apps enable personalized delivery of such interventions while collecting behavioral feedback from travelers. However, designing interventions personalized to individual travelers becomes increasingly complex when accounting for time-varying network traffic dynamics. For example, the optimal intervention for a traveler may depend not only on their characteristics and preferences but also on the evolving congestion patterns and system-wide travel behavior. To address this, the behavioral intervention design is formulated as a time-varying problem to adapt to varying traffic dynamics. A reinforcement learning (RL) framework integrated with the SUMO traffic simulator is proposed to simulate the interactions of urban traffic environment, traveler heterogeneity, behavioral response, and mode choices over time. The RL agent learns optimal intervention strategies, such as the informational messages to provide to subgroups, or monetary interventions to provide on a network link by learning the traveler preferences and traffic dynamics of the network in real time. The framework offers a scalable tool for traffic operators to design and deploy interventions based on behavioral feedback and traffic dynamics.

3 - A Simulation-Optimization Framework and Web-Based Application for Dynamic Volunteer Routing in Food Banking Systems

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

Efficient and equitable delivery in not-for-profit food logistics is challenged by dynamic customer demand and fluctuating volunteer availability. This study introduces a simulation-optimization framework to support real-time volunteer routing under uncertainty. The objective is to minimize overall delivery costs while maintaining operational feasibility across dynamically evolving scenarios. The model captures decentralized operational constraints and incorporates stochasticity in both customer arrivals and volunteer participation. A Monte Carlo Tree Search (MCTS) algorithm is embedded in a discrete-event simulation to iteratively update delivery routes based on emerging requests, aiming to minimize delivery costs while maintaining operational feasibility.

Comparative analyses show that MCTS outperforms traditional methods such as greedy insertion and fixed-horizon value iteration, offering significant improvements in cost efficiency by anticipating future demand and adapting routes in real time.

To enable real-world deployment, a web-based application is developed and integrated with the optimization engine. The platform connects households, food pantries, and volunteers through a unified interface, supporting request submission, inventory management, volunteer scheduling, and real-time route dispatch. This integrated system ensures coordinated, efficient, and equitable food delivery across the community.

4 - Real-World Exposure to Last-Mile Delivery Externalities: A Novel Approach Using Vehicle and Human Trajectory Data

Franco Basso, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile, Sara Perotti, Raúl Pezoa, Francisco Fuentes, Julio Covarrubia

Research on last-mile delivery services and their environmental externalities has traditionally focused on emissions at the source, with limited attention to how these effects are distributed across urban populations. However, negative externalities such as pollution and noise do not impact all residents equally, as exposure is shaped by spatial and temporal proximity to the source. This paper proposes a novel framework to measure the heterogeneous impact of last-mile delivery by integrating high-resolution GPS data from vehicles with dynamic demographic information derived from mobile phone traces. Using econometric models—including zero-inflated count and linear regression techniques—we quantify how delivery vehicle activity correlates with population density and sociodemographic characteristics throughout the day in Santiago, Chile. Our findings reveal substantial disparities in exposure across age, gender, and socioeconomic groups, underscoring the uneven burden of urban logistics on public health and environmental justice. This framework enables more targeted and equitable mitigation strategies for urban freight operations by identifying who is most affected and where. The results highlight the need for logistics planning that accounts not only for operational efficiency but also for its differentiated impact on urban communities.

5 - Optimizing Fire Station Placement for Growing Urban Demand: a Case Study in Goldsboro, North Carolina

Sarah Halsey, North Carolina State University, Raleigh, NC, United States, Demetra Protoggyrou, Mallory Padgett, Catherine Reckard, Leila Hajibabai

This study applies facility location models to support fire station planning for the Goldsboro Fire Department (GFD) in North Carolina. Two mathematical models were used to guide decision-making: a set-covering model to determine the minimum number of stations required to cover all demand points, and a maximum coverage model to identify optimal station placement under resource constraints. The models incorporate both current incident data and projected growth patterns, and apply distance-based constraints (e.g., 1.5-mile service radius) as a proxy for time-based response standards.

The set-covering model successfully generated a configuration that met full coverage requirements, including anticipated growth in the city's northern regions. However, the solution required opening more stations than the city's budget could support. To address this, the maximum coverage model was used to optimize coverage given a limited number of stations. This model revealed inefficiencies in the current network—specifically, overlapping service areas among several stations—and recommended relocating the headquarters station to improve spatial coverage without increasing the number of facilities.

Goldsboro has seen significant residential development in its northwest and northeast corridors in recent years, resulting in increased incident call volume and challenges in meeting national fire response standards. These modeling efforts provide GFD with actionable insights on how to restructure its station layout to maintain service quality in the face of growth and limited resources. This case highlights how location modeling can support infrastructure planning and enhance resilience in small and mid-sized urban settings.

6 - A Novel Approach to Two-Way Arterial Traffic Signal Coordination

JinWoo Na, Pohang University of Science and Technology, Nam-Gu, Pohang-Si, Korea, Republic of, Jun Kim, Indrayoga Putra, Byung-In Kim

Traffic signal coordination plays a critical role in managing traffic flow across urban arterial networks under varied traffic conditions. Selecting an appropriate signal coordination strategy suitable for the traffic conditions is essential for minimizing delays and enhancing traffic-flow efficiency. Maximum bandwidth-based forward progression such as MAXBAND, MULTIBAND, and AM-BAND have been devised and applied to coordinate two-way arterial traffic signals in undersaturated, near-saturated, or moderately congested traffic conditions. MAXBAND establishes uniform progression bands across the artery but fails to consider segment-specific traffic conditions. MULTIBAND addresses this limitation by introducing variable-bandwidth progression in each segment. However, its symmetrical bandwidth allocation restricts the full utilization of available green time. By contrast, AM-BAND facilitates asymmetrical bandwidth allocation but exhibits high sensitivity to the Q-value, which governs the relative distribution of the left and right sides of the bandwidth. Determining an optimal Q-value for varying traffic conditions is further complicated by the absence of a well-defined mathematical relationship between Q-value and performance. Therefore, this study proposes a novel approach which integrates the uniform band structure of MAXBAND with segment-specific asymmetrical bandwidth allocation while addressing their respective limitations. The approach maintains progression efficiency across the artery while providing flexibility to accommodate varying traffic demands. The proposed model was formulated as a mixed-integer linear programming model and evaluated through SUMO traffic simulation in a realistic environment, where all arterial intersections are controlled by a pre-timed strategy with a common cycle length. Experimental simulation results demonstrate that the proposed model outperforms existing progression models.

7 - Monotonic Neural Networks for Crash Severity Prediction at Highway-Rail Grade Crossings

Amin Keramati, Widener University, Chester, PA, United States, Pan Lu

Highway-Rail Grade Crossings (HRGCs) are spatial locations where highways and rail tracks intersect at the same grade, posing a risk of collisions between trains and vehicles. Due to (1) the inability of trains to stop promptly after braking and (2) the substantial mass difference between trains and road users, crashes at HRGCs often result in severe outcomes. Consequently, transportation agencies and decision-makers must implement effective safety strategies to reduce crash severity, typically categorized into three levels: Property Damage Only (PDO), Injury, and Fatal.

A widely adopted approach in the literature for supporting safety decision-making at HRGCs involves developing predictive models that assess the influence of various crash factors on the likelihood and severity of incidents. However, most existing models estimate the likelihood of each crash severity level independently, without accounting for potential dependencies among them. Furthermore, few studies have applied advanced machine learning or statistical models to capture the dynamic impact of crash factors on severity and its variation over time.

In this study, we propose a novel deep learning framework—**Monotonic Neural Network (MNN)**—for time-to-severity event analysis at HRGCs. The model captures the interdependence among the three severity levels, aligning with the inherent characteristics of HRGC crashes. We design and evaluate two MNN architectures: **Multi-Head**, and **Single-Head**. Comparative analysis shows that the best-performing architecture improves prediction accuracy by **3% to 30%**, demonstrating the potential of our approach to support proactive safety interventions.

WA19

Building A Level 4 A405

Learning and Control of Stochastic Systems

Invited Session

Applied Probability Society

Chair: Yue Hu, Stanford University, Stanford, CA, United States

Co-Chair: Vahid Sarhangian, University of Toronto, Toronto, ON, Canada

1 - Learning to Defer in Content Moderation: the Human-AI Interplay

Wentao Weng, MIT, Cambridge, MA, United States, Thodoris Lykouris

Successful content moderation in online platforms relies on a human-AI collaboration approach. A typical heuristic predicts the expected harmfulness of a post and uses fixed thresholds to decide whether to remove it and whether to send it for human review. This disregards the prediction uncertainty, the time-varying element of human review capacity and post arrivals, and the selective sampling in the dataset (humans only review posts filtered by the admission algorithm).

We introduce a model to capture this human-AI interplay. Our algorithm observes contextual information for posts, makes classification and admission decisions, and schedules them for human review. Only admitted posts receive human reviews on their harmfulness. These reviews help educate the machine-learning algorithms but are delayed due to congestion in the human review system. We propose a near-optimal learning algorithm that balances the classification loss from a selectively sampled dataset, the idiosyncratic loss of non-reviewed posts, and the delay loss of having congestion in the human review system.

Our results offer concrete insights on how online platforms can harness the power of AI while ensuring system reliability. First, congestion-aware admission rules are necessary to handle the time-varying nature of online requests and human capacity. Second, popular optimism-only online learning heuristics (used in content moderation practice) can fail to learn capable AI models as they disregard label collection that is necessary for future classification decisions. To the best of our knowledge, this is the first result for online learning in contextual queueing systems, which may be of independent interest.

2 - Joint Learning and Pricing in Many-Server Queues: near-Optimal Policies via Fluid Duality

Xu Zheng, University of Miami, Coral Gables, FL, United States

We study the problem of joint learning and dynamic pricing in a large-scale service system modeled as a many-server queue, where customer arrival rates depend on posted prices through an unknown demand function. The system manager seeks to maximize expected profit—defined as revenue minus congestion costs—by learning the demand function and integrating it into pricing decisions. We propose a novel explore-cooling-exploit policy that integrates structured exploration with state-dependent pricing control. Our approach leverages a fluid approximation and dual-based reformulation to design lightweight, closed-loop pricing rules that respond to real-time system conditions. We analyze the regret of this policy under both finite and continuous pricing settings within a many-server asymptotic framework. Our theoretical results establish tight regret bounds, revealing an economy-of-scale effect where learning becomes more efficient as the system size grows. Numerical experiments validate the effectiveness of our policy and highlight the distinct trade-offs between exploration and congestion in large-scale service environments. This work is the first to provide provably near-optimal closed-loop joint learning and pricing policies for many-server queues under demand uncertainty.

3 - Online Learning for Dynamic Service Mode Control

Wenqian Xing, Stanford University, Stanford, CA, United States, Yue Hu, Anand Kalvit, Vahid Sarhangian

Modern service systems are increasingly adopting innovative modalities enabled by emerging technologies, such as AI-assisted services, to achieve a better balance between quality and efficiency. Motivated by these advancements, we study the optimal dynamic control of service modes in systems with unknown parameters, aiming to balance reward accumulation and system congestion. We consider a single-server queueing system with two switchable service modes, each characterized by a distinct

service rate and an unknown distribution of rewards earned upon service completion. The objective is to maximize the long-run average of cumulative rewards minus holding costs. To address this problem, we first characterize the optimal state-dependent policy under full knowledge of model parameters. For the case of unknown reward distributions, we propose an online learning algorithm based on Upper Confidence Bound estimates to adaptively learn the optimal policy. Our algorithm achieves a statistically near-optimal regret bound of the square root of T over a time horizon T and demonstrates strong numerical performance. A key methodological contribution is a novel regret decomposition and a regenerative cycle-based framework, offering broader insights into learning-based optimal control in queueing systems. Finally, we demonstrate the practical relevance of our approach through a case study on optimizing AI-assisted patient message replies in a healthcare setting.

4 - A Planning Framework for Adaptive Labeling

Daksh Mittal, Columbia University, New York, NY, United States, Yuanzhe Ma, Shalmali Joshi, Hongseok Namkoong

Ground truth labels/outcomes are critical for advancing scientific and engineering applications, e.g., evaluating the treatment effect of an intervention or performance of a predictive model. Since randomly sampling inputs for labeling can be prohibitively expensive, we introduce an adaptive labeling framework where measurement effort can be reallocated in batches. We formulate this problem as a Markov decision process where posterior beliefs evolve over time as batches of labels are collected (state transition), and batches (actions) are chosen to minimize uncertainty at the end of data collection. We design a computational framework that is agnostic to different uncertainty quantification approaches including those based on deep learning, and allows a diverse array of policy gradient approaches by relying on continuous policy parameterizations. On real and synthetic datasets, we demonstrate even a one-step lookahead policy can substantially outperform common adaptive labeling heuristics, highlighting the virtue of planning. On the methodological side, we note that standard REINFORCE-style policy gradient estimators can suffer high variance since they rely only on zeroth order information. We propose a direct backpropagation-based approach, Smoothed-Autodiff, based on a carefully smoothed version of the original non-differentiable MDP. Our method enjoys low variance at the price of introducing bias, and we theoretically and empirically show that this trade-off can be favorable.

5 - Optimal Noise Schedules for Diffusion Models

Billy Jin, Purdue University, West Lafayette, IN, United States, Baris Ata, Harsha Honnappa

Diffusion models have emerged as a cornerstone of modern generative AI, achieving state-of-the-art performance in image, audio, and video generation. However, training these models involves numerous design choices and hyperparameters that can significantly impact performance. In this talk, I will present a principled framework for selecting the noise schedule, providing theoretical foundations that justify commonly used heuristics.

WA20

Building A Level 4 A406

Topics in Control, Learning and Optimization of Stochastic Systems

Invited Session

Applied Probability Society

Chair: Prakash Chakraborty, The Pennsylvania State University, State College, PA, United States

1 - Optimizing Telemedicine Referrals to Mitigate ED Overcrowding

Sabha Baniyadi, The Pennsylvania State University, State College, PA, United States, Prakash Chakraborty, Paul Griffin

Emergency department (ED) overcrowding in the US has become a critical public health issue leading to increased wait-times. This delay compromises patient care, strains resources, and increases the risk of adverse outcomes. In this research, we mitigate congestion by identifying an optimal threshold policy for telemedicine referrals of non-urgent patients in a hospital emergency department while simultaneously maximizing revenue generation and minimizing waiting and balking costs. We assume a two-class M/M/c priority queueing system, with different service time distributions for urgent and non-urgent patients. Subsequently, we develop a threshold-based policy to reduce patient congestion and waiting times, thereby enhancing overall revenue generation for the hospital. This approach provides hospital administrators with an implementable decision framework for efficient ED management.

2 - Many Server Functional Fluid Limit for Non-Stationary Loss Models

Mingrui Wang, Penn State University, state college, PA, United States, Prakash Chakraborty

Modern service systems often operate under non-stationary demand and non-Markovian dynamics with limited server and buffer capacities. Classical stationary models fail to capture the dynamic and constrained nature of these environments, motivating the analysis of non-stationary, non-Markovian loss queueing systems. In this research, we study such systems in the many-server regime and establish a functional strong law of large numbers (FSLN) for three core performance processes: the fraction of occupied servers, the fraction of occupied buffers, and the departure rate. These coupled processes are shown to converge to the unique solution of a system of Volterra integral equations. Numerical experiments demonstrate the accuracy of the fluid approximation in estimating blocking probabilities, highlighting its potential as a scalable performance analysis tool.

3 - Exponentially Shifted Gaussian Smoothing for Decision-Dependent Optimization Problems

Prakash Chakraborty, The Pennsylvania State University, State College, PA, United States, Mingrui Wang, Uday Shanbhag

Decision-dependent optimization problems—where the objective or constraints are influenced by the decision variables themselves—arise frequently in machine learning, operations research, and other domains. Addressing these challenges without gradient information necessitates efficient zeroth-order (ZO) methods. However, existing ZO techniques often suffer from poor scalability due to unfavorable dependence on problem dimension. To overcome this, we propose an Exponentially-Shifted Gaussian Smoothing (esGS) estimator, developed via a simple change-of-variable approach. The esGS estimator achieves tighter moment bounds and significantly mitigates dimensional sensitivity. When incorporated into a ZO optimization framework, it leads to improved iteration complexity with rigorous convergence guarantees for both convex and strongly convex settings. Preliminary experiments demonstrate its practical effectiveness, particularly in decision-dependent scenarios.

4 - Stochastic Games with Information Cost in Directed Chain Network

Tomoyuki Ichiba, University of California-Santa Barbara, Santa Barbara, CA, United States

We shall consider a class of stochastic differential games of infinitely many players in a directed chain network with partial information. We discuss the convergence of the Nash equilibria from finitely many players to infinitely many players. Some explicit solution is obtained under the linear quadratic problem and then the cost of acquiring additional information is explored.

WA21

Building A Level 4 A407

Advances in Urban and On-Demand Delivery Systems

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Xinwei Chen, The University of Tulsa, Tulsa, OK, United States

1 - Analyzing Fulfillment Dynamics in on-Demand Grocery Delivery Systems

Alp Arslan, Lancaster University, Lancaster, United Kingdom, Jelder van der Gaast, Jan Fransoo

The rapid growth of on-demand delivery services (ODS) has transformed customers' online grocery shopping expectations. ODS offer increased convenience and speed through micro-fulfillment centers or partner-retail stores. Determining the most effective ODS fulfillment strategy across market settings, such as the presence or absence

of micro-fulfillment centers, remains an open question. Additionally, the utility of technology acquisition, such as investing in systems that monitor product availability across retail stores to reduce operational uncertainty, requires further investigation. We study this problem by developing a stylized queueing system that allows us to analytically assess an ODS's performance under a variety of configuration strategies. Our model captures crucial ODS characteristics, including stochastic order arrivals, uncertain product availability, stochastic picking times, and tight service level requirements. We provide explicit expressions for calculating the service time and click-to-delivery time, which depend on factors such as assortment and stock-out uncertainties. These expressions facilitate comparisons between various ODS configurations. Our findings suggest that micro-fulfillment centers may outperform models relying on partner-retail stores in scenarios where assortment information is not shared, in-stock availability is relatively low, and average order basket sizes are substantial. However, when a certain level of information sharing is introduced, incorporating assortment and stock-in data from stores, employing partner-retail stores may prove more efficient in terms of factors like service time and courier utilization compared to relying solely on micro-fulfillment centers.

2 - Unlocking Comodal Rural Mobility Through Integrated Passenger and Package Delivery Services

Xinwei Chen, The University of Tulsa, Tulsa, OK, United States, Qi Luo

Rural and suburban communities are largely overlooked in the design of affordable and reliable on-demand mobility solutions due to sparsely distributed populations, the prevalence of long-distance travel needs, and limited access to alternative transit services. In this research, we propose a comodal mobility system that integrates passenger transport and package delivery, ensuring seamless coordination between the two compatible services within a single network. Our research is comprised of (1) analytical insights into the feasibility conditions for integrating on-demand transit with package delivery, and (2) an online optimization framework for real-time order dispatching in comodal scenarios. Case studies demonstrate the potential of comodality strategies to expand access to essential services, particularly for carless households and individuals with mobility limitations.

3 - A General Decomposition Framework for Dynamic Two-Stage Order Fulfillment Problems

Gal Neria, Tel Aviv University, Tel Aviv, Israel, Michal Tzur, Marlin Ulmer

Coordinating fulfillment operations across two interdependent stages is a critical challenge across industries such as logistics, e-commerce, and food services. In this work, we introduce the Dynamic Two-Stage Order Fulfillment Problem (DTS-OFPP), a new class of problems that unifies several logistical challenges in a cohesive manner. The DTS-OFPP captures the complexities of real-world fulfillment systems and emphasizes the need to simultaneously optimize both preparation and delivery stages under dynamic and stochastic conditions.

We model the problem class as Markov Decision Process and introduce a novel theory-driven policy decomposition formulation.

To exploit this decomposition, we propose an innovative solution framework that integrates operations research elements and machine learning.

Our framework advances beyond prior research, which often treats fulfillment stages independently or sequentially, and offers a new approach to handling interdependencies between sequential operations.

Our framework is applied to three real-world DTS-OFPP contexts: warehouse picking operations, production-delivery systems, and food delivery platforms. Numerical experiments using datasets from Meituan, ORTEC, and Iowa City show that our framework consistently outperforms existing methods.

4 - Who Benefits from Drone-Enhanced Food Delivery? An Economic Analysis of the Food Delivery Market

Chaoyu Wu, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China, People's Republic of, Xiaotong Sun, chenyang li

Amid growing interest in drone-enhanced food delivery, most studies have focused on operational challenges, leaving the strategic and economic implications understudied. This talk addresses that gap by developing a parsimonious economic model of on-demand instant delivery market that introduces drones alongside human couriers, capturing strategic interactions between consumers, couriers, and the platform. The model employs a linear city framework with spatially distributed consumers with heterogeneous delivery valuations. Key factors are endogenized, including drone efficiency, pricing, and courier supply. We compare outcomes under three market structures (perfect competition, monopoly, and Bertrand duopoly) across three delivery modes: human-only, drone-only, and bi-modal. Analysis reveals that drone adoption expands delivery coverage beyond the baseline human-only mode and is most cost-effective for long-distance orders. In the bi-modal setting, non-overlapping service regions emerge due to rational consumer choice, yielding maximum social welfare across all market structures. Market structure also influences pricing: under duopoly competition, the introduction of drones drives down delivery fees for remote customers. However, increasing drone penetration diminishes the surplus of human couriers, thereby raising equity concerns for the workforce. Numerical simulations with real data validate our findings, demonstrating that integrating drones significantly expands delivery coverage and enhances customer surplus across all market structures, thereby positioning customers as the primary beneficiaries. These findings provide managerial insights for platform strategy and highlight policy interventions, such as drone subsidies or profit redistribution, to ensure the efficiency gains from drone adoption do not come at the expense of equity.

WA22

Building A Level 4 A408

Service and Logistics Operations Interface

Invited Session

Applied Probability Society

Chair: Daisik Nam, Inha University, Michul-gu, N/A, Korea, Republic of

Co-Chair: EunSu Lee, New Jersey City University, Jersey City, NJ, United States

1 - Leveraging Generative AI for Logistics Data Analysis and Decision-Making

Daisik Nam, Inha University, Incheon, Korea, Republic of, Jaeyoung Lee, Minyoung Park

This study presents a practical framework for implementing generative artificial intelligence (GenAI) in logistics operations, focusing on automated reporting and data analysis workflows. As logistics organizations face increasing demands for operational efficiency and data-driven decision making, systematic approaches to GenAI integration become essential for competitive advantage.

We demonstrate a hands-on implementation approach using Cursor AI for logistics workflow automation, specifically targeting automated inventory reporting and shipment analysis. The methodology involves a structured four-step process: data upload, natural language command execution, result analysis, and iterative improvement through prompt refinement. Through practical case studies using real logistics data, we show how organizations can achieve significant time savings in routine reporting tasks while maintaining accuracy and consistency.

The framework addresses key application areas including automated inventory summaries, shipment delay pattern analysis, and operational dashboard generation. Implementation results demonstrate measurable improvements in reporting efficiency, with case studies showing substantial reductions in manual data processing time. The study provides practical guidelines for logistics practitioners to implement GenAI tools effectively while addressing common challenges such as data accuracy verification and workflow integration.

This research contributes actionable implementation strategies for logistics managers seeking to leverage GenAI for operational automation, emphasizing practical deployment approaches rather than theoretical frameworks.

2 - Developing a Screening Methodology for Dedicated Freight Truck Lanes Using Transportation Big Data and Traffic Simulation

Daejin Kim, Inha University, Incheon, Korea, Republic of

This study proposes a methodology for identifying candidate sections for dedicated freight truck lanes by integrating big transportation data with traffic simulation techniques. Utilizing real-world traffic volume data from the Korea Transport Institute's View-T system, the analysis evaluates traffic and environmental factors across the national road network. Thirteen key indicators—including freight vehicle ratio, average speed, congestion intensity, and pollutant emissions—were selected, normalized

using Min-Max scaling, and aggregated into a composite Screening Score. Based on this score, high-priority road segments were systematically identified. The prioritization process was further refined by incorporating road attributes such as road classification, with weighted adjustments reflecting freight traffic volume to better capture freight-dense corridors. To assess the potential impacts of implementing dedicated freight lanes, top-ranked sections underwent microscopic traffic simulation analysis using VISSIM, evaluating metrics such as changes in average speed, congestion relief, emissions reduction, and traffic safety improvements. This study presents a practical and data-driven analytical framework that supports both strategic screening and simulation-based policy assessment for freight infrastructure planning.

3 - Quantifying the Benefits of Cargo Bikes for Last-Mile Delivery: Operational, Environmental, and Social Perspectives

Sungmin Kim, Inha University, Incheon, Korea, Republic of, Gwanyong Gu, Daejin Kim

The rapid growth of e-commerce has substantially increased urban freight traffic, intensifying challenges such as congestion, environmental pollution, and traffic-related safety risks. In South Korea, for instance, parcel volumes have risen at an average annual rate of 12.53%, further worsening these issues. Conventional delivery trucks are often restricted from entering residential complexes, alleyways, and traditional markets, leading to frequent illegal parking and traffic obstruction. As a sustainable and space-efficient alternative, cargo bikes have emerged as a promising last-mile delivery solution in dense urban areas.

This study assesses the operational, environmental, and safety benefits of cargo bikes through a comparative field experiment conducted over a two- to three-week period in a designated regulatory zone in Gimcheon, South Korea. The area includes both dense residential neighborhoods and modern apartment complexes. Two scenarios were compared: the AS-IS scenario, where diesel trucks made individual deliveries, and the TO-BE scenario, where cargo bikes performed consolidated deliveries from a shared depot in a nearby parking facility. GPS tracking and delivery logs were used to evaluate efficiency in terms of delivery volume, travel time, and distance.

Results show that about 3.3 cargo bikes can replace one conventional delivery truck in terms of capacity. This substitution is projected to reduce annual last-mile operating costs by over USD 7,500 and significantly lower accident-related social costs and environmental impacts. These findings support integrating cargo bikes into urban logistics strategies—especially in areas with limited truck access—while enhancing efficiency, safety, and sustainability.

4 - A Unified Framework for Container Drayage Optimization: Integrating Task Selection with Operational Constraints

Dongyun Kim, Graduate School of Logistics, Incheon National University, Incheon, Korea, Republic of, Kwangsup Shin, Yunhong Min

Drayage is the short-haul trucking of containers between intermodal hubs—ports, rail yards, warehouses, and customer sites—covering the first- or last-mile of a container's journey. Its effectiveness depends on precise scheduling and routing of trucks and containers to meet time windows and maximize efficiency. This study proposes an integrated drayage model that enhances the efficiency and practicality of empty container utilization by embedding task selection into the planning process. Each customer request for an empty container can be satisfied by choosing one of several related tasks that reflect operational constraints such as live operations, dual transactions at terminals, and terminal-container compatibility. The model jointly schedules and routes trucks to minimize both the number of unsatisfied requests and individual truck completion times. Computational experiments are performed to demonstrate the effectiveness of the proposed model and solution.

5 - Innovative Process and Strategic Management for Enhancing Shipbuilding Capacity

EunSu Lee, New Jersey City University, Jersey City, NJ, United States, Luis Alberto, Eun Soo Kim

This presentation examines the current status and production capacity of the U.S. shipbuilding industry and identifies the key challenges limiting its competitiveness in the global market. It explores a range of innovative manufacturing processes and advanced strategic management approaches designed to strengthen operational efficiency, optimize resource utilization, and expand overall shipbuilding capacity.

WA24

Building A Level 4 A410

Scalable Optimization for Large Foundation Models and DNNs

Invited Session

OPT: Nonlinear Optimization

Chair: Yuchen Lou, Northwestern University, Evanston, IL, United States

Co-Chair: Minshuo Chen, Northwestern University, Evanston, IL, United States

1 - Modelling Compositionality and Tensor Structure in Neural Network Optimization

Jeremy Bernstein, MIT, Cambridge, MA, United States

An old idea in optimization theory says that since the gradient is a dual vector it may not be subtracted from the weights without first being mapped to the primal space where the weights reside. In this talk, we take this idea seriously and construct a duality map for general neural networks. Our map, which we call modular dualization, forms a unifying theoretical basis for training algorithms that are a) fast and b) scalable. Modular dualization involves first assigning operator norms to layers based on the semantics of each layer, and then using these layerwise norms to recursively induce a duality map on the weight space of the full neural architecture. We conclude by deriving GPU-friendly algorithms for dualizing Embed, Linear and Conv2D layers -- the latter two methods are based on a rectangular Newton-Schulz iteration (Kovarik, 1970; Björck & Bowie, 1971). A variant of our methods was used to set speed records for training NanoGPT. Overall, we hope that our theory of modular duality will yield a next generation of fast and scalable optimizers for general neural architectures.

2 - Reimaging Gradient Descent: Large Stepsize, Oscillation, and Acceleration

Jingfeng Wu, University of California, Berkeley, Berkeley, CA, United States

Gradient descent (GD) and its variants are pivotal in machine learning. Conventional wisdom suggests smaller stepsizes for stability, yet in practice, larger stepsizes often yield faster convergence, despite initial instability. In this talk, I will explain how large stepsizes provably accelerate GD for logistic regression on separable data in three settings.

- We start with logistic regression, a convex but non-strongly convex problem. We show that GD with a large stepsize attains an ϵ -risk in $O(1/\sqrt{\epsilon})$ steps. This matches the accelerated step complexity of Nesterov momentum and improves the classical $O(1/\epsilon)$ step complexity for GD with a small stepsize.
- We then consider ℓ_2 -regularized logistic regression with regularization strength λ , a strongly convex problem with condition number $\Theta(1/\lambda)$. We show GD with a large stepsize attains an ϵ -excess risk in $O(1/\sqrt{\lambda} \ln(1/\epsilon))$ steps. This, again, matches Nesterov momentum and improves the $O(1/\lambda \ln(1/\epsilon))$ step complexity for GD with a small stepsize.
- Finally, we consider the task of finding a linear separator of a linearly separable dataset with margin γ . We show that, with large and adaptive stepsizes, GD solves this task in $1/\gamma^2$ steps by minimizing the logistic risk. We further show this step complexity is minimax optimal for all first-order methods, and cannot be achieved by GD with small stepsizes.

The talk is based on a sequence of papers in collaboration with Peter Bartlett, Matus Telgarsky, Bin Yu, Ruiqi Zhang, Licong Lin, and Pierre Marion.

3 - Purifying Shampoo: Understanding and Improving Shampoo Through An Approximation Perspective

Anna Cai, Meta, San Mateo, CA, United States, Runa Eschenhagen

The recent success of Shampoo in the AlgoPerf contest has sparked renewed interest in Kronecker-factorization-based optimization algorithms for training neural networks. Despite its success, Shampoo relies heavily on several heuristics such as learning rate grafting and stale preconditioning to achieve performance at-scale. These

heuristics increase algorithmic complexity, necessitate further hyperparameter tuning, and lack theoretical justification. This paper investigates these heuristics from the angle of Frobenius norm approximation to full-matrix Adam and decouples the preconditioner's eigenvalues and eigenbasis updates. We show that grafting from Adam mitigates the staleness and mis-scaling of the preconditioner's *eigenvalues* and how correcting the eigenvalues directly can eliminate the need for learning rate grafting. To manage the error induced by infrequent *eigenbasis* computations, we propose an adaptive criterion for determining the eigenbasis computation frequency motivated by terminating a warm-started QR algorithm. This criterion decouples the update frequency of different preconditioner matrices and enables us to investigate the impact of approximation error on convergence. These practical techniques offer a principled angle towards removing Shampoo's heuristics and developing improved Kronecker-factorization-based training algorithms.

4 - Recent Advances in Kronecker-Factored Preconditioned Stochastic Gradient Methods for Large-Scale Training

Hao-Jun Shi, Meta Platforms, Inc., Menlo Park, CA, United States

Kronecker-factored preconditioned stochastic gradient methods, such as Shampoo and Muon, offer a more sample-efficient alternative to diagonal-scaling-based methods like Adam and AdaGrad for training recommendation systems and large language models (LLMs). Notably, Muon, a stochastic version of the steepest descent method under the spectral norm, can be interpreted as a special case of Shampoo. This talk further explores this relationship through a formal decomposition of Shampoo as a covariance-adapted version of Muon. In order to better understand the effect of different terms in the decomposition, we delve into the role of epsilon, pseudo-inverses, and numerics in Shampoo and Muon, and present detailed ablations that systematically incorporate the covariance adaptation term. Furthermore, we address the challenge of scaling these methods to large-scale models that require parameter sharding via Fully Sharded Data Parallel (FSDP) and other parallelisms, presenting both shard-dependent and shard-independent approaches to overcome scalability issues. By shedding light on the connections between Shampoo and Muon, and providing insights into their scalability, this talk aims to contribute to our understanding of both the systems and training efficiency tradeoffs of different non-diagonal optimization techniques when applied to structured large-scale machine learning problems with different noise and curvature properties.

5 - Structured Positive-Definite Curvature Learning

Wu Lin, vector institute, Toronto, ON, Canada

Stochastic positive-definite curvature estimation plays a crucial role in machine learning. For example, Adam(W) estimates a diagonal curvature matrix for neural network training. Other notable applications include covariance estimation in Bayesian deep learning and gradient-free evolutionary optimization. Leveraging the structure of curvature matrices—such as Kronecker-based, non-diagonal structures—is essential for moving beyond diagonal approximations in large-scale settings. However, existing structured curvature estimation methods, such as Shampoo, are often specialized for particular types of curvature information, limiting their adaptability in addressing challenges posed by non-standard applications. To address this, we propose a Riemannian optimization approach to structured curvature estimation that supports flexible forms of curvature information. We demonstrate the efficacy and versatility of our method in positive-definite matrix optimization and covariance adaptation for gradient-free optimization, as well as its efficiency in curvature learning for neural network training.

WA25

Building A Level 4 A411

Learning and Simulation: Interplay and Advances

Invited Session

Simulation

Chair: Karthyek Murthy, University of Southern California, Los Angeles, CA, 90089, United States

1 - Efficient Gradient Estimation for Overparameterized Stochastic Systems

Shengbo Wang, University of Southern California, Los Angeles, CA, United States, Jose Blanchet, Peter Glynn

Overparameterized dynamic stochastic models have achieved remarkable success in various complex environments, such as stochastic control, reinforcement learning, financial engineering, and machine learning systems. These models often feature system evolution that is parameterized by a high-dimensional vector $\theta \in \mathbb{R}^n$, aiming to optimize a path-dependent expectation such as a value function through stochastic gradient ascent. Consequently, designing efficient gradient estimators for which the computational complexity scales well with n is of significant interest. We introduce a novel unbiased stochastic gradient estimator—the generator gradient estimator—for which the computation time remains stable in n . In addition to establishing the validity of our methodology for stochastic differential equations and Markov jump processes, we also perform numerical experiments, demonstrating a significant improvement in efficiency compared to the widely used pathwise differentiation method: our estimator achieves near-constant computation times and increasingly outperforms its counterpart as n increases. We also generalize our approach to policy optimization in a reinforcement learning environment where the underlying system dynamics are unknown.

2 - Cheap Iterated Bootstrap for Higher-Order-Accurate Uncertainty Quantification

Yunhao Yan, Columbia University, New York, NY, United States, Shengyi He, Henry Lam

The bootstrap is a versatile method for quantifying statistical uncertainty. Among its variants, a popular approach, the studentized bootstrap, provably achieves higher-order coverage error reduction compared to other benchmarks. However, without further structural knowledge, the latter implementation requires iterated resampling which can make the computation expensive, especially for large-scale models such as those arising in stochastic simulation. In this talk, we introduce a "cheap" iterated bootstrap method that substantially thins one of the sampling layers in running the studentized bootstrap, via an integrated view of the uncertainty coming from the resample distribution and the Monte Carlo computation in implementing the resampling. Intriguingly, in building our methodology, we also show the intimate relation of the studentized bootstrap with the studentized t-distribution, presenting a distinct view from the common claim of the informality of "studentization" in the bootstrap literature.

3 - Learning Optimal Solutions to Chance-Constrained Optimization Models from Limited Data

Karthyek Murthy, University of Southern California, Los Angeles, CA, United States

We consider the challenge of provisioning resources to meet high reliability or service level requirements, which are typical in applications such as power distribution, emergency response, and regulatory risk management. With chance-constrained optimization serving as a natural starting point for modeling this class of problems, our main contribution is to characterize how the optimal costs and decisions scale for a generic joint chance-constrained model as the target probability of satisfying the reliability constraints approaches its maximal level. Besides providing insights into the behaviour of optimal costs and decisions based on large-deviations theory, the scaling result carries several algorithmic implications, beginning with the following observation on distributionally robust optimization (DRO) modeling of chance constraints: While widely used DRO approaches based on KL-divergences, Wasserstein distances, and moments distort the scaling properties of optimal decisions and lead to exponentially higher costs, incorporating marginal distributions or employing suitably chosen f -divergence balls preserves the correct scaling and ensures that decisions remain conservative by at most a constant or logarithmic factor. Further, given N data samples, we demonstrate how the scaling can serve as a basis for estimating approximately Pareto-optimal decisions whose constraint violation probabilities can even be smaller than the $\Omega(1/N)$ —estimation barrier that is fundamental in the absence of parametric assumptions.

4 - The Scale Invariance of Portfolio Risk Measurement

Nifei Lin, Laboratory for AI-Powered Financial Technologies Limited, Hong Kong, China, People's Republic of, Jeff Hong

The risk measurement of derivative portfolios is typically formulated as a nested estimation problem. The traditional studies on such problem, including standard nested simulation and smoothing approaches, focus on the convergence rate of the estimator with respect to the sample size of portfolio losses. In practice, however, the core computational burden arises from simulating risk factors and subsequently repricing each instrument in the portfolio, which grows with both the number of instruments and the dimensionality of the risk factors. While the classical analysis framework is adequate for small- and medium-scale portfolios, it becomes increasingly inefficient

for large-scale portfolios, whose computation cost is expensive and performance may suffer from curse of dimensionality. To address the challenges posed by large-scale portfolios, we revisit the standard problem formulation and highlight a key but underutilized structural feature: the additive nature of portfolio losses, which allows us to group multiple instruments in the portfolio with the same underlying risk factors. Leveraging this property, we propose independent sampling and one-hot sampling methods to deal with the repricing for each group of instruments, which are compatible with existing nested estimation frameworks. We prove that, by integrating portfolio scale into the asymptotic analysis, our methods yield a scale-invariant property: the overall computational cost, measured in terms of computation time rather than sample size, remains in the same order of magnitude regardless of portfolio size. In other words, the convergence rate is independent with the scale of the portfolio.

WA26

Building A Level 4 A412

Generative AI for Decision-Making and Optimization

Invited Session

Simulation

Chair: Giulia Pedrielli, Arizona State University, Tempe, AZ, United States

Co-Chair: Raul Astudillo, California Institute of Technology, Pasadena, CA, United States

Co-Chair: Matthias Poloczek, Amazon, San Francisco, CA, United States

1 - Llms for Instance-Specific Cutting Plane Separator Configuration

Yingxi Li, Stanford University, Stanford, CA, United States, Connor Lawless, Anders Wikum, Madeleine Udell, Ellen Vitercik

Mixed integer linear programming (MILP) solvers come with numerous parameters that are difficult to select a priori for all but expert users, yet they significantly impact solver performance. Current machine learning (ML) methods for configuring solvers require training models on thousands of MILP instances, generalize poorly to new problem sizes, and often demand complex ML pipelines and custom solver interfaces, making integration into existing workflows challenging. In this paper, we introduce an LLM-based framework for configuring cutting plane separators for a given MILP problem with little to no training data. Our pipeline can be applied to MILPs even without prior knowledge of the problem. We augment these LLMs with descriptions of cutting plane separators available in a given solver, grounded by summarizing the existing research literature on separators. While individual solver configurations have a large variance in performance, we present an ensembling strategy that clusters and aggregates configurations to create a small portfolio of high-performing configurations. Our LLM-based methodology requires no custom solver interface, can find high-performing configurations by solving only a few MILPs, and can generate the configuration with simple API calls that run in under a second. Numerical results show our approach is competitive with existing configuration approaches on a suite of classic combinatorial optimization problems and real-world MILPs with only a fraction of the training data and computation time.

2 - Extracting Knowledge Priors with LLMs for Bayesian Optimization

Jacob Gardner, University of Pennsylvania, Philadelphia, PA, United States

Algorithmic advances in the efficiency, scalability, and applicability of Bayesian optimization have made it competitive or even state-of-the-art on some of the most fascinating optimization problems of our time. Recent advances in language modeling have led to an unprecedented ability to retrieve and digest natural language text, and reason about the content of that text. Why not use these models to featurize this text in ways that, for example, a Gaussian process surrogate model can consume during optimization? In this talk, I'll present ongoing work where we are trying to make the knowledge stored in text accessible to Bayesian optimization. As one concrete example, we will consider the problem of augmenting commonly used in silico molecule design benchmarks with a new mandate: the designed molecules must not be toxic or mutagenic. We show that, by fine-tuning LLMs to tag molecules across hundreds of millions of paragraphs found in scientific papers and then by building CLIP embeddings, we can build zero-shot molecular property predictors that, in some cases, achieve more than 90% test precision despite never seeing labeled data. By combining this with Bayesian optimization's relatively newfound ability to handle large numbers of constraints, we achieve a(n early!) prototype of a pretty cool promise: we can simply write down in English properties we'd like a molecule or protein to have, and then optimize a molecule or protein subject to the constraint that it has those properties.

3 - AI and the Future of Optimization Modeling

Madeleine Udell, Stanford University, Stanford, CA, United States, Ali AhmadiTeshnizi, Wenzhi Gao, Connor Lawless

Optimization problems are pervasive in sectors from manufacturing and distribution to healthcare. However, most such problems are still solved heuristically by hand rather than optimally by state-of-the-art solvers, as the expertise required to formulate and solve these problems limits the widespread adoption of optimization tools and techniques.

As a glimpse of the future, this talk will introduce OptiMUS, a Large Language Model (LLM)-based agent designed to formulate and solve MILP problems from natural language descriptions.

OptiMUS can develop mathematical models, write and debug solver code, develop tests, and check the validity of generated solutions.

Experimentally, OptiMUS correctly solves more than 80% of benchmark problems, more than twice as many as a basic LLM prompting strategy.

More broadly, we discuss the potential for LLMs in domains where accuracy and fidelity to real-world data is critical and strategies to augment and safeguard their performance.

4 - Direct Regret Optimization in Bayesian Optimization

Fengxue Zhang, University of Chicago, Chicago, IL, United States, Yuxin Chen

Bayesian optimization (BO) is a powerful paradigm for optimizing expensive black-box functions. Traditional BO methods typically rely on separate hand-crafted acquisition functions and surrogate models for the underlying function, and often operate in a myopic manner. In this paper, we propose a novel **direct regret optimization** approach that jointly learns the optimal model and non-myopic acquisition by distilling from a set of candidate models and acquisitions, and explicitly targets minimizing the multi-step regret. Our framework leverages an **ensemble of Gaussian Processes (GPs)** with varying hyperparameters to generate simulated BO trajectories, each guided by an acquisition function chosen from a pool of conventional choices, until a **Bayesian early stop criterion** is met. These simulated trajectories, capturing multi-step exploration strategies, are used to train an end-to-end decision transformer that directly learns to select next query points aimed at improving the ultimate objective. We further adopt a **dense training--sparse learning** paradigm: The decision transformer is trained offline with abundant simulated data sampled from ensemble GPs and acquisitions, while a limited number of real evaluations refine the GPs online. Experimental results on synthetic and real-world benchmarks suggest that our method consistently outperforms BO baselines, achieving lower simple regret and demonstrating more robust exploration in high-dimensional or noisy settings.

WA41

Building B Level 3 B301

Healthcare Innovation

Invited Session

Health Applications Society

Chair: Dan Yamin, Tel Aviv University Iby and Aladar Fleischman Faculty of Engineering, P.O. Box 39040, Tel Aviv 6997801 1, Rehovot, 69978, Israel

1 - Terminating Pandemics with Smartwatches

Martial Ndeffo, Texas A&M University, College Station, TX, United States, Märt Vesinurm, Dan Yamin, Margaret Brandeau

Studies have demonstrated that wearable devices, such as smartwatches, can accurately detect infections in presymptomatic and asymptomatic individuals. Yet, the extent to which smartwatches can contribute to prevention and control of infectious diseases through a subsequent reduction in social contacts is not fully understood. We developed a multiscale modeling framework that integrates within-host viral dynamics and between-host interactions to estimate the risk of viral disease outbreaks within a given population. We used the model to evaluate the population-level effectiveness of smartwatch detection in reducing the transmission of three COVID-19 variants and seasonal and pandemic influenza. With a 66% reduction in contacts after smartwatch-based disease detection, we estimate that the reproduction number R would drop from 2.55 (interquartile range [IQR]: 2.09–2.97) to 1.37 (IQR: 1.00–1.55) for the ancestral COVID-19 variant; from 1.54 (IQR: 1.41–1.69) to 0.82 (IQR: 0.68–0.85) for the delta variant; from 4.15 (IQR: 3.38–4.91) to 2.20 (IQR: 1.57–2.52) for the omicron variant; from 1.55 (IQR: 1.34–1.74) to 0.81 (IQR: 0.63–0.87) for pandemic influenza; and from 1.28 (IQR: 1.18–1.35) to 0.74 (IQR: 0.64–0.79) for seasonal influenza. With a 75% reduction in contacts, R decreases below 1 for the delta variant and for pandemic and seasonal influenza. Sensitivity analyses across a wide array of parameter values confirm that self-isolation initiated shortly after smartwatch detection could significantly reduce R under diverse epidemiological conditions, different levels of smartwatch detection accuracy, and realistic self-isolation levels. Our study underscores the revolutionary potential of smartwatches to manage seasonal diseases and alter the course of future pandemics.

2 - Multi-Scale Modeling of Periodicity for Missing Data Imputation in TIME Series: a Smartwatch Application

Yuewei Ling, Stanford University, Stanford, CA, United States, Edan Shahmoon, Margaret Brandeau, Dan Yamin

Missing data is a challenge in healthcare time-series analysis. Periodic patterns influence both the observations and missingness mechanism, particularly in wearable technologies that produce rich physiological signals, but most current methods ignore these underlying cycles. We introduce a Periodic Multi-scale Imputation (PMSI) framework to impute missing data, which models multi-scale periodic dependencies for both observations and missingness patterns. The imputation process follows three stages: (1) identify candidate periodicities; (2) transform series into multi-dimensional tensors aligned with selected periodicities; (3) optimize hyperparameters using a two-layer minimization of Akaike information criterion (imputation parameters for each dimension and dimensionality). PMSI is applied first to model the missingness mechanism and generate synthetic masks to observations for their optimizations. We validated PMSI using a large smartwatch-derived heart rate dataset, consisting of 1,674 participant-month series in a high-adherence cohort (smartwatch wearing adherence) and 5,871 series in a low-adherence cohort. Spectral analysis revealed daily and 3.5-day (likely smartwatch charging) periodicities in the missingness mechanism, and daily and weekly periodicities in observations. The optimal case used hourly resolution and daily periodicity to impute missing values for PMSI, and outperformed five standard imputation methods: k-nearest neighbors, autoregressive integrated moving average, convolutional neural networks, piecewise linear interpolation, and last observation carried forward. PMSI yielded the lowest RMSE: 6.92 versus 7.74–9.59 in the high-adherence cohort, and 8.65 versus 9.30–14.54 in the low-adherence cohort (all adjusted $p < 0.01$). Our PMSI framework enhances imputation accuracy effectively by modeling periodic dependencies, and holds great promise to impute any time-series with periodicities in healthcare.

3 - Self-Reported Recovery Versus Smartwatch-Detected Recovery for COVID-19, Influenza and Group A Streptococcus: a Two-Year Prospective Cohort Study

Varun Gande, Stanford University, Stanford, CA, United States, Dan Yamin, Margaret Brandeau, Erez Shmueli

Timely and accurate detection of recovery from infectious diseases is essential for minimizing complications, optimizing care, and preventing long-term health issues. This study analyzes recovery dynamics for COVID-19, influenza, and group A streptococcus (GAS) using data from a two-year prospective cohort of 4,795 participants in Israel. We integrate smartwatch-derived physiological and behavioral data with self-reported symptoms and electronic medical records to assess recovery at three key points: (1) end of moderate-to-severe symptoms, (2) full symptom resolution, and (3) return to physiological baseline—termed “digital recovery.” Results reveal a consistent lag between self-reported recovery and physiological normalization, with longer delays observed in moderate-to-severe cases. For COVID-19, digital recovery occurred a median of 7.1 days post-symptom resolution in mild cases, and 60.2 days in moderate-to-severe cases. Influenza showed lags of 2.5 and 7.9 days, and GAS showed 1.1 and 4.4 days, respectively. Notably, daily step counts and reported activities returned to baseline once symptoms resolved, suggesting behavioral recovery precedes physiological recovery. These findings challenge current public health guidelines that recommend return to normal activities five days post-symptom cessation. Our results indicate longer recovery windows—up to two months for severe COVID-19 cases—may be warranted. This framework highlights the value of integrating wearable sensor data into infectious disease monitoring, offering scalable tools for recovery assessment, intervention planning, and policy optimization.

WA42

Building B Level 3 B302

Equity in Health Policy: Shaping the Future of Fair and Informed Care

Invited Session

Health Applications Society

Chair: Amaya McNealey, Georgia Institute of Technology, Atlanta, GA, United States

1 - Mind the Gap: Methods to Evaluate the Quality of Clinical Care

Divya Shanmugam, Cornell Tech, NYC, NY, United States

The evaluation of clinical care is fundamentally constrained by the need for large, labeled datasets—resources that range from expensive to impossible to obtain. This talk introduces methods to evaluate the quality of clinical care by making use of unlabeled data to evaluate decision-making of both humans and algorithms. In the first part, I focus on retrospective evaluation of past diagnostic decisions, and present a framework to measure the rate of missed diagnoses in the health record. In the second part, I turn to prospective evaluation of algorithmic decision-making in clinical care, and introduce a method to assess the performance of clinical predictive models in the absence of abundant labeled data. Together, these projects highlight a shared challenge in both human and algorithmic decision-making—evaluating decision quality under data constraints—and provide steps towards robust evaluation of both human and algorithmic care.

2 - Modelling the Cost-Effectiveness of a Combined Rsv Prevention Strategy

Leonardo Basso, Instituto Sistemas Complejos de Ingeniería, Santiago, Chile

Recent developments in RSV prevention—including the availability of the RSVpreF maternal vaccine, nirsevimab, and clesrovimab—present an important opportunity to model the potential benefits of combined immunisation strategies. Performing simulation analyses on the integrated national databases that we have been using, we propose to assess the cost-effectiveness and expected health impact of various implementation scenarios that integrate these three Immunoprophylactic agents, considering variations in their mixture within a universal prevention strategy. Our focus will be on estimating reductions in RSV-related hospitalisations, medical leaves, and on computing performance metrics such as the number needed to immunise (NNI). Key parameters of the modelling will be efficacy and/or effectiveness, costs and expected coverage of each monoclonal; these values will be allowed to change within reasonable boundaries. It is possible to create a visual tool to analyze different cases easily. This modelling effort will provide timely and actionable evidence to support strategic planning by health authorities in Chile and other countries with similar demographic and epidemiological profiles

3 - Estimating the Effectiveness of Nirsevimab in High-Risk Infant Populations

Denis Saure, Universidad de Chile, Santiago, Chile, Leonardo Basso, Marcel Goic, Charles Thraves, Miguel O'ryan, Juan Torres, Felipe del Solar, Ignasi Neira, Amal Zgheib

Respiratory syncytial virus (RSV) is a leading cause of infant hospitalization, especially among preterm infants and those with congenital heart disease (CHD). This study evaluates the impact of Chile's nationwide Nirsevimab immunization program on RSV-related hospitalizations in high-risk infant populations.

We conducted a matched case-control study using national administrative and clinical data covering over 145,000 immunized infants. RSV-related hospitalizations were identified by linking ICD-10 discharge codes to confirmed RSV cases through optimization-based calibration using sentinel surveillance data. This approach allowed for precise classification of RSV-related lower respiratory tract disease (LRTD) events in the absence of universal virologic testing.

To control for confounding, we employed a mathematical integer programming model to match each RSV-related case to seven controls based on infant age, hospital admission date, and geographic region. This 1:7 matching scheme optimized statistical power and balance across key covariates.

Effectiveness of the Nirsevimab immunization was estimated using the matched cohort. Further, we applied simulation modeling to estimate the overall reduction in RSV-related hospitalizations attributable to the immunization strategy in the high-risk population.

Findings demonstrate that the targeted Nirsevimab campaign substantially reduced RSV-related LRTD hospitalizations among preterm and CHD infants. This study underscores the value of combining optimization and simulation methods for evaluating public health interventions and informing resource allocation decisions.

WA43

Building B Level 3 B303

Emerging Issues at the Interface of Supply Chain and Finance

Invited Session

MSOM: iForm

Chair: M. Hakan Hekimoglu, Rensselaer Polytechnic Institute, Troy, NY, United States

1 - On the Benefits of Supplier Vs Customer Diversification: a Geographical Analysis

Deepak Agrawal, Carnegie Mellon University, Pittsburgh, PA, United States, Nikolay Osadchiy

Diversification is traditionally considered a risk mitigation tool. In our paper, we examine both supplier and customer diversification, geographically, and find that diversification also have first order effect on inventory productivity. We find a nonlinear relationship between geographical diversification and inventory turnover. Further, we examine the varying effects of diversification by firm's upstreamness.

2 - Internal Credit and External Blame: Self-Attribution In Operations and Supply Chain Performance

Yimeng Niu, Tilburg University, Tilburg, Netherlands, Xiao Qiao, Jing Wu, Xingsheng Yang

Problem definition: This paper examines self-attribution tendencies within corporate settings, specifically in the context of operations and supply chain performance. Managers often attribute superior company performance to internal factors, such as product innovation and intra-organizational teamwork ("internal credit"), while attributing poor performance to external factors, such as supply chain partners or macroeconomic conditions ("external blame"). **Methodology/Results:** We fine-tune large language models (LLMs) to analyze extensive question-and-answer transcripts from earnings conference calls, developing a measure of company-level self-attribution. Through empirical investigations, we identify a consistent pattern of self-attribution, characterized by internal credit and external blame, in discussions related to operations and supply chain issues. Our findings indicate that companies with larger sizes and more experienced managers exhibit stronger self-attribution tendencies. Furthermore, firms with lower operational efficiencies and supply chain risks are more likely to display this behavior. This self-attribution, on average, seems healthy and improving. Companies with stronger self-attribution tendencies are more likely to optimize their internal operations, enhance cash flow and production efficiency, as well as make strategic supply chain adjustments. These firms also demonstrate a greater ability to coordinate effectively with customers, replace underperforming suppliers, and retain high-performing ones in the subsequent quarter. **Managerial implications:** Managers can leverage self-attribution tendencies as a cost-effective operational risk-hedging tool to safeguard their reputation and better position their organizations to proactively address operational challenges and adjust to poor performance outcomes

3 - Vintech: Investing in Wine Using Analytics

M. Hakan Hekimoglu, Rensselaer Polytechnic Institute, Troy, NY, United States, Burak Kazaz

This paper examines a comprehensive path regarding how analytics can be employed for automated trading of wine. We coin this as VinTech. The paper highlights the essential components of data analytics necessary for robo-advising. We first develop a pricing algorithm to represent the realistic value of a wine. Using this algorithm, we then compare the price evolution of underpriced wines to overpriced wines. Last, our robo-advising algorithm constructs wine portfolios for different investment goals and risk preferences. We demonstrate the performance of our algorithm using out-ofsample testing. This paper aims to convert an opaque market into a transparent and efficient investment market.

WA44

Building B Level 3 B304

Sustainable Supply Chain Management

Contributed Session

Contributed

Chair: Akshay Mutha, University of Vermont, Burlington, VT, United States

1 - Dual-Channel Supply Chain Contract Design Considering Retailer Greening Cost

Yuan Yuan Deng, Huazhong University of Science and Technology, Wuhan, China, People's Republic of, Haijun Wang

In dual-channel supply chain, the retailer's possession of private cost information gives rise to information asymmetry, which may lower supply chain efficiency and affect supply chain coordination. Based on dual-channel supply chain consisting of a manufacturer and a retailer, this paper analyzes the design of revenue-sharing contract considering the retailer's green investment cost under both full and asymmetric information, and explores the impact of the retailer's green investment level on the decisions in the dual-channel supply chain. The result shows that the impact of the retailer's green investment level on the supply chain members' optimal decisions may be affected by information asymmetry. In addition, we explore the impact of cost information asymmetry on the performance of dual-channel supply chain by comparing the optimal decisions under the two cases and analyze the value of information. We show that the retailer's asymmetric cost information hurts the manufacturer's profit and the efficiency of the whole supply chain, but it weakly benefits the retailer.

2 - A Stochastic Optimization Model for Consumer-Responsive Sustainable Supply Chain Management Under Market Uncertainty

Ikhyun Jo, Yonsei University, Seoul, Korea, Republic of, Soongeol Kwon

As consumer interest in sustainability continues to grow, companies increasingly recognize the importance of integrating consumer preferences into their sustainable supply chain management (SSCM) strategies. These preferences, as reflected in purchasing behavior, are shaped by companies' long-term and consistent sustainability efforts, and their relative importance may vary depending on the specific dimensions of sustainability under consideration. In this study, we propose an optimization model for SSCM that incorporates consumers' long-term and differentiated preferences across various sustainability factors. To quantitatively capture these preferences and analyze their implications, we develop a mixed-integer linear programming (MILP) model. The model features consumer-responsive demand that evolves over time based on the supply chain's sustainability performance. Additionally, the uncertain proportion of green consumers significantly influences SSCM decision-making. To handle this uncertainty, we use a two-stage stochastic programming approach, generating time-series scenarios for the proportion of green consumers. Through numerical experiments, we evaluate the proposed consumer-responsive SSCM model and demonstrate the critical importance of understanding the proportion of green consumers and their purchasing behavior in shaping effective sustainability strategies.

3 - More or Less: Climate Pledges and Supply Chain Flexibility

Jie Lian, The University of Oklahoma, Norman, OK, United States, Keith Skowronski, Yan Dong

Firms increasingly make environmental commitments, including climate pledges, which can significantly influence the operational decisions of their suppliers. For example, suppliers may reduce inventory buffers and supply diversification in response to decreased uncertainty in products and technologies. This study focuses on the supplier decision-making process and examines how climate pledges affect key operational choices that shape supply chain flexibility.

4 - Tracing Influence: ESG Spillovers in Complex Supply Chain Networks

Francis Bilson Darku, University of Notre Dame, Notre Dame, IN, United States, Rick Hardcopf, Sarvanan Devaraj

As environmental, social, and governance (ESG) practices become central to corporate strategy, firms face increasing scrutiny and evaluation from stakeholders and ESG rating agencies. While substantial research exists on the drivers of ESG adoption and abandonment, less is known about how ESG behaviors diffuse across interconnected firms, particularly within complex supply chain networks. This study explores the extent to which a focal firm's ESG ranking is shaped by the ESG performance of its suppliers, customers, and industry peers. Leveraging longitudinal data from Bloomberg's SPLC and ESG datasets, we analyze over 1,600 U.S. manufacturing firms and their supply chain partners over a nine-year period. We identify key actors within these networks and quantify their influence on ESG performance, offering new insights into the dynamics of ESG practices. Our findings aim to inform both scholars and practitioners about the relational drivers of ESG performance in supply chains.

5 - Product Acquisition for Remanufacturing

Akshay Mutha, University of Vermont, Burlington, VT, United States, Zhaolin Li, Daewon Sun, Jennifer Ryan

We develop models to analyze the process of acquiring used products for remanufacturing. We perform numerical analyses to show the applicability of our models.

WA46

Building B Level 3 B306

The Expanding Role of GenAI in Data-Driven Decisions in Digital Society

Invited Session

Decision Analysis Society

Chair: Dongtan Li, Auburn University, Auburn, AL, 36830, United States

1 - Exploring the Impact of Genai Integration on Behavioral Shifts in Multifaceted Online Communities

Dongtan Li, Auburn University, Opelika, AL, United States

With the advent of generative AI (GenAI), online platforms navigate critical decisions regarding its integration, as GenAI has reshaped user experiences and behaviors in digital society. To compete with external GenAI influence, many platforms embed GenAI algorithms into their services to enhance user engagement. This study investigates how GenAI integration affects user behavior within multifaceted online communities (MOC), which facilitate both knowledge exchange and social communication. By leveraging data from Reddit, we examine changes in user information seeking and sharing patterns after implementing Reddit Answers. Additionally, we explore how these behavioral shifts are moderated by forum characteristics, differentiating between social-oriented and knowledge-oriented subreddits. Our analysis employs a Difference-in-Differences estimation with propensity score matching to provide robust results. Findings reveal significant alterations in posting and commenting behaviors that vary by forum type. These insights offer valuable guidance for platform designers and community managers seeking to optimize AI-enhanced interactions.

2 - The Impact of ESG Risks on Product Recalls

Niyanta Sookraj, Auburn University, Auburn, AL, United States, Yen-Yao Wang

The growing prominence of environmental, social, and governance (ESG) concerns has drawn increased attention from investors, regulators, and corporate leaders alike, positioning ESG risk management as a central force in shaping business strategy. This shift toward sustainability is emphasized by substantial financial commitments, with billions of dollars being directed toward ESG-related initiatives. Prior research highlights that firms failing to manage ESG risks are more vulnerable to financial and reputational damage. However, beyond these external consequences, ESG risks may also pose considerable operational threats. While digital tools for sustainability reporting and ESG monitoring are increasingly adopted, empirical research linking ESG risks to operational failures remains scarce.

This study addresses that gap by examining the relationship between ESG risks and product recall events, with a focus on how ESG risks interact with consumer feedback to offer a more holistic view of recall dynamics. Since poor ESG performance would erode consumer trust and increase the likelihood of adverse market reactions, understanding the role consumers play in this dynamic is essential. Leveraging artificial intelligence (AI) to assess consumer sentiment, we demonstrate how advanced information systems can enhance firms' capabilities to detect, assess, and mitigate operational vulnerabilities. Drawing on 2,894 manufacturer-month observations from 22 U.S. automotive firms over a ten-year span, we find that elevated ESG risk ratings significantly increase the likelihood of product recalls. Additionally, consumer complaints amplify this effect, underscoring the interaction between ESG risk exposure and stakeholder responses.

3 - Domino Effects of AI: Spillovers from New App Launches on Developer Portfolio

Yalda Hojjati, Auburn University, Auburn, AL, United States, Niyanta Sookraj, Pankush Kalgotra

Technological innovations often disrupt and reshape digital markets by altering how value is created and consumed. Such disruptions may trigger shifts in the usage of related products and services within the ecosystem. This study explores these intra-platform spillovers through the lens of artificial intelligence (AI)—a highly disruptive and rapidly diffusing technology. Mobile apps represent a key digital platform where AI is increasingly embedded, as they serve as one of the fastest-growing and most direct channels for companies to integrate AI into consumer-facing services. This makes understanding AI's impact on app ecosystems critical. We focus on a key question: How does the launch of a new app—AI or non-AI—affect user engagement with a developer's existing apps? Further, how do these spillover effects differ based on the number of AI and non-AI apps already in the developer's portfolio? To answer these questions, we used data mining techniques to construct a novel longitudinal dataset from the Apple App Store, capturing biweekly observations over six months. Using two-way fixed effects panel regression, we find that new app launches generally boost engagement with existing apps. However, the magnitude and direction of this effect depend on the portfolio's AI composition. Launching a new AI app increases engagement more when the existing portfolio is primarily non-AI, while this effect diminishes in AI-heavy portfolios. Conversely, launching a non-AI app reduces engagement when more AI apps already exist. The findings offer strategic value for AI integration and advance theoretical understanding of demand spillovers and innovation-led disruption in digital platforms.

4 - Impact of Store Openings and Closures on Local Retail Networks: a Community Detection Approach

Amir Mousavi, Auburn University, Auburn, AL, United States, Pankush Kalgotra

This study investigates the impact of store openings and closures on overlapping communities within retail networks. As retail landscapes continuously evolve through dynamic entry and exit patterns, understanding how these changes affect existing market structures becomes crucial for business strategy. We employ the Democratic Estimate of the Modular Organization of a Network (DEMON) algorithm to detect overlapping communities within retail networks, capturing the complex interdependencies that characterize modern retail ecosystems (Coscia et al., 2012).

We examine how opening and closure events affect the structure and performance of retail networks. DEMON's community detection capabilities enable identification of overlapping communities where retail establishments participate in multiple market segments simultaneously. This approach reflects the reality that stores often serve diverse consumer groups across different spatial and commercial boundaries. Our methodology analyzes how retail network topology changes following market disruptions and examines the spillover effects on neighboring competitors.

<p break-words="">Building on literature examining dynamic market evolution and spatial retail interactions (Shen, 2014; Sung, 2022; Fang & Yang, 2024), we explore how competitive dynamics influence retail network resilience. The study contributes to understanding how localized openings and closures impact interconnected retail systems via overlapping communities. This network-based perspective provides insights into the cascading effects of retail transformations, offering valuable implications for stakeholders regarding market stability, competitive positioning, and the sustainability of local retail ecosystems in the face of continuous market changes.

WA47

Building B Level 3 B307

Optimization Topics III

Contributed Session

Contributed

Chair: Marçal Ferreira, Professor Marçal, Rua Itacolomi, 180, São Paulo, 01239020, Brazil

1 - Bilevel Optimization with Non-Convex Lower-Level Problem

Xiaotian Jiang, University of Minnesota, Minneapolis, MN, United States, Ioannis Tsaknakis, Jiaxiang Li, Jiawen Bi, Prashant Khanduri, Mingyi Hong, Shuzhong Zhang

Bilevel optimization (BLO) problem, where two optimization problems (referred to as upper- and lower-level problems) are coupled hierarchically, has wide applications in areas such as machine learning and operations research. We investigate bilevel optimization problems where the lower-level problem is nonconvex. To address the inherent difficulty of such settings, we propose two distinct reformulations. The first method focuses on unconstrained and nonconvex lower-level problems and introduces a new hyperfunction defined via accumulation points induced by a fixed algorithm and initial point for lower-level problem. Since the resulting hyperfunction is typically discontinuous, we apply a smoothing technique and solve its smoothed approximation by sampling the upper-level variable to estimate its gradient. The second method targets constrained and nonconvex lower-level problems by approximating the value function through sampling feasible solutions, leading to a convex surrogate. This surrogate is used to construct a penalized reformulation of the original BLO problem. For both methods, we develop gradient-based algorithms with provable convergence guarantees and validate their effectiveness through numerical experiments.

2 - ODH for FICO® Xpress: Redefining the Limits of Optimization

Robert Ashford, Optimization Direct, Inc., Thurlaston, United Kingdom, Alkiviadis Vazacopoulos

As optimization challenges grow increasingly complex across industries, traditional solvers often fall short when faced with large-scale, multi-constraint decision models. **Optimization Direct Heuristics (ODH) for FICO® Xpress** represents a transformative enhancement in optimization technology—enabling organizations to overcome long-standing performance barriers without replacing existing systems. This paper introduces the breakthrough capabilities of ODH, highlighting its unique ability to complement FICO Xpress by intelligently combining advanced heuristics, structural decomposition, and parallel processing. Rigorous testing demonstrates ODH's impact: faster solve times, significant reductions in optimality gaps, and the ability to find feasible solutions where other methods fail. With demonstrated success in domains such as telecommunications, retail supply chains, energy generation, and healthcare scheduling, ODH is not just a technical innovation—it is a strategic enabler of real-time, high-quality decision-making. This work outlines the technical architecture, implementation strategies, and quantifiable ROI of ODH, establishing it as a pivotal advancement for organizations striving to transform operational complexity into competitive advantage.

3 - Using Knapsack Problems to Motivate Fortified Regression

Hugo Mainguy, University of Maryland, College Park, MD, United States, Bruce Golden, Luca Bertazzi

The knapsack problem can be computationally expensive to solve to optimality. Although it is generally easier to solve than many other NP-Hard problems, there are many instances that remain computationally expensive to solve. In these circumstances, it may be useful to estimate the optimal solution values. In this paper, we develop a simple variant of linear regression and test it by appealing to the knapsack problem literature. We take advantage of well-known knapsack problem datasets in order to illustrate the potential benefits of fortified regression.

4 - A Taxonomy of Control Problems in Boolean Networks

Kyungduk Moon, Pohang University of Science and Technology, Pohang, Korea, Republic of, Céilia Biane, Kangbok Lee, Loïc Paulevé

Boolean networks are discrete dynamical models that use Boolean states and logical transition formulas to capture the dynamic behavior of a system. Identifying a set of controls—defined as modifications of transition formulas—is crucial to ensure that a desired condition is satisfied across all target states. However, existing computational tools often produce inconsistent solutions due to differences in modelling assumptions. We propose a taxonomy that explains these inconsistencies and clarifies relationships among the tools. This framework is demonstrated with Python code that systematically compares them.

5 - Multi-Timescale Distributed Communication Sliding for Distributed Optimization

Junhui Zhang, MIT, Cambridge, MA, United States, Patrick Jaillet

We present first-order methods for convex non-smooth distributed optimization problems with the special structure that the saddle point formulation is dual-block-decomposable. Such structure arises when the consensus constraints are decomposed into blocks, which are penalized separately. We propose a set of conditions for these penalty functions under which duality gaps for the saddle point problem provide upper bounds on objective value suboptimality and consensus constraint violation for the original distributed optimization problem.

Taking advantage of the block structure, we propose the multi-timescale distributed communication sliding algorithms, where dual blocks are updated at different frequencies. We show that the duality gap converges at the rate of $\mathcal{O}(\overline{r}/k)$ for Lipschitz objectives, and $\mathcal{O}(C(\overline{r}, \overline{r^2})/\mu k^2)$ for μ -strongly convex objectives, where $\overline{r}$ and $\overline{r^2}$ measure "average (squared) rate of updates" for dual blocks, and $C(\overline{r}, \overline{r^2})$ depends on $\overline{r}$ and $\overline{r^2}$.

The deterministic block updating rules and the flexibility in picking the updating frequencies allow the design of more efficient algorithms, especially when the costs of updating different duals are different. Moreover, with proper choices of the penalties, the communication round complexities of our MT-DCS achieve optimal dependency on function similarity, a more refined measure of the variation between the local objectives than Lipschitz constants.

6 - Efficient Wave Planning in Omnichannel Retail: Minimizing Sku Repetition with Random Keys Optimizer

Marçal Ferreira, ITA / UNIFESP, São Paulo, Brazil, Luiz Leduino, Antônio Chaves, Guilherme Coelho

In omnichannel retail, repeated SKU picks across production waves inflate labor costs and slow throughput. We address this problem in partnership with Lojas Renner S.A., one of Brazil's largest fashion chains, minimizing SKU repetition while respecting capacity and homogeneity constraints. We cast it as a graph-partitioning problem: boxes are vertices, and shared SKUs induce edges whose separation across waves incurs cost. We apply the Random-Key Optimizer (RKO) framework to encode box-to-wave assignments as continuous random keys. Our decoder enforces (i) homogeneity of wave class and stock type; (ii) capacity limits for boxed and hanging stock. Unlike existing greedy heuristics, RKO guarantees both capacity compliance and class homogeneity in large-scale instances. On real instances, RKO reduces SKU repetition versus a best-fit heuristic, while never increasing — and often reducing — the number of waves. The results demonstrate that RKO delivers a scalable, high-quality decision-support tool for improving picking efficiency and operational throughput in large-scale, multi-wave production systems.

WA48

Building B Level 3 B308

Machine Learning for Better Healthcare Decisions

Invited Session

Health Applications Society

Chair: Yifang Yan, The Pennsylvania State University, University Park, PA, United States

Co-Chair: Qiushi Chen, The Pennsylvania State University, University Park, PA, United States

1 - Robust AI-Assisted Radiation Therapy Planning

Arka Roy, The University of Texas at San Antonio, San Antonio, TX, United States, Ruiqi Li, Nikos Papanikolaou

Artificial intelligence (AI)-based radiation therapy planning offers the potential to enhance treatment efficiency and expand access to care, particularly in underserved and rural communities facing workforce shortages. However, the inherent errors in AI-generated treatment plans can compromise clinical effectiveness. As the adoption of AI in clinical practice accelerates, there is a critical need for methods that can quantify and mitigate model-related errors. In response, we propose an integrated predictive-prescriptive modeling framework that fuses the contouring and treatment planning stages of radiation therapy. This approach is designed to produce treatment plans that are robust to segmentation inaccuracies, ultimately improving the reliability and safety of AI-driven care.

2 - A Machine Learning Approach to Facilitate Optimal Allocation of Hard-to-Place Kidneys

Sait Tunc, Virginia Tech, Blacksburg, VA, United States, Sean Berry, Berk Gorgulu, Mucabit Cevik

The growing concerns surrounding the organ allocation systems, particularly in light of persistently high nonuse rates for kidneys, has intensified calls for more transparent, data-driven methods. Public concern over ad hoc, opaque decision-making has highlighted the limitations of existing policies, especially when it comes to allocating hard-to-place kidneys (e.g., from older donors, with prolonged ischemia, or donor comorbidities). We present a unified machine learning framework designed to improve the placement of such kidneys. We develop interpretable models to predict transplant center acceptance behavior and propose a ranking-based intervention system that prioritizes centers most likely to utilize the offered organ. Our results demonstrate a significant reduction in delays and nonuse rates. We will also discuss ongoing work on optimizing the batch size of simultaneously expiring organ offers. This effort balances the risks of sequential offers, leading to delays and nonuse, against the inefficiencies of overly large batches, which can overwhelm transplant centers and diminish future offer credibility. Throughout this work, interpretability remains central: our aim is not to replace decision-makers but to support them with robust, transparent tools that adapt to evolving clinical practices.

3 - Neural Index Policies for Restless Multi-Action Bandits with Heterogeneous Budgets

Himadri Pandey, Georgia Institute of Technology, Atlanta, GA, United States, Gian Garcia, Kai Wang

Restless multi-armed bandits (RMABs) with multiple actions and heterogeneous budget constraints pose a major challenge for sequential decision-making under uncertainty, with broad applications in areas such as healthcare and resource allocation. Existing RMAB approaches, including Whittle index policies, are typically restricted to binary actions and struggle to adapt to complex budget constraints. We propose a novel `neural index policy` that assigns learnable indices to multiple actions and uses a knapsack formulation to translate these indices into a feasible action policy under multiple constraints. To enable end-to-end training, we reformulate the knapsack selection step as a differentiable optimal transport (OT) problem and solve it using the Sinkhorn algorithm, allowing gradients to propagate through the selection process. The neural network is trained by minimizing the discrepancy between its induced occupancy measure and the optimal occupancy measure obtained from a linear programming relaxation. Our approach enables scalable and effective policy learning for multi-action RMABs under complex constraints, with experimental results demonstrating strong approximation.

4 - Learning from Simulated Patient Trajectories: A Sequential Decision-Making Agent for Heart Transplantation

Yingtao Luo, Carnegie Mellon University, Pittsburgh, PA, United States, Reza Skandari, Carlos Martinez, Arman Kilic, Rema Padman

Decisions about managing patients on the heart transplant waitlist are currently made by committees of doctors who consider multiple factors, but the process remains largely ad-hoc. With the growing volume of longitudinal patient, donor, and organ data collected by the United Network for Organ Sharing (UNOS) since 2018, there is increasing interest in analytical approaches to support clinical decision-making at the time of organ availability.

We introduce an agentic decision-making framework based on model-based reinforcement learning (MBRL) for optimizing long-horizon, sequential decisions in heart transplantation. Our approach leverages time-to-event models for waitlist mortality, transplant mortality, and donor offer arrival to simulate full patient trajectories. These predictive models serve as environment dynamics, enabling the agent to learn policies that make timely, personalized decisions to maximize patient outcomes.

By capturing temporal dependencies and clinical trade-offs across the transplant process, our framework supports decision-making beyond static risk models or myopic heuristics. The learned policy identifies key predictors aligned with known clinical risk factors while surfacing novel insights, offering a foundation for urgency assessment and policy refinement in heart transplant care. The agent supports realistic counterfactual reasoning, allowing us to evaluate what might happen under alternative action sequences for individual patients. It also captures the temporal dependencies, uncertainty, and trade-offs inherent in the transplant process—such as urgency, donor-recipient matching, and long-term outcomes—through full patient trajectory simulation. This work illustrates the potential of agentic AI systems to support transparent, consistent, and patient-centered clinical decision-making in high-stakes settings.

5 - Making Multi-Class Prediction Actionable: An Application to Reduce Missed Primary Care Appointments

Yifang Yan, The Pennsylvania State University, University Park, PA, United States, Qiushi Chen

Missed primary-care appointments disrupt continuity of care, leading to poorer health outcomes, higher healthcare costs, and wasted clinical resources. Previous studies have developed risk prediction models for missed appointments to understand the patients' no-show behavior and to guide patient outreach strategies. However, most studies focus on predicting binary outcomes, rarely distinguishing late cancellations from no shows, despite their different impacts on clinical workflow in practice. For multi-class prediction problems, prediction results are typically presented as a set of predicted probabilities, making it difficult to inform actions directly. To bridge these gaps, we present a novel classification with inconclusiveness framework based on a disjunctive optimization model, which translates predicted probabilities into deterministic classes and can also identify inconclusive cases with low predictive confidence. The optimization model minimizes the misclassification loss weighted by potential impact on patients' health outcomes. We demonstrate this approach in the context of primary care appointments for adults with type 2 diabetes. We use real-world patient health records from 15 family medicine clinics at a regional academic health center to predict three classes (no-show, late cancellation, and completed visit) and solve for the optimal classification rule to determine confident and inconclusive cases. In our case study, we develop a simulation model of a patient navigation center with limited patient outreach resources to evaluate the potential health impact of various patient classification rules. Our optimal classification rule yields fewer QALYs lost attributed to misclassifications compared with strategies that prioritize by no-show risk alone or by potential health impact alone.

WA49

Building B Level 3 B309

Financial Adaptation and Resource Allocation under Climate Risk

Invited Session

ENRE: Energy-Climate

Chair: Felipe Verastegui, Columbia University, New York, NY, United States

Co-Chair: Agostino Capponi, Columbia University, New York, NY, United States

1 - Market Mechanisms to Improve Efficiency with Green Energy Resources

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

Adapting to changing climate conditions will require integration of extensive renewable resources. Current energy markets have innate inefficiencies that can lead to significant increases in costs of adaptation. This talk will discuss the use of expanded markets with contingent claims to reduce or eliminate these inefficiencies and to enable a smooth energy transition.

2 - Innovator's Edge In Supply Chain Transparency and Food Waste Reduction

Chenghuai Li, University of Alberta, Edmonton, AB, Canada, Bora Keskin, Jing-Sheng Jeannette Song

New digital technologies are transforming fresh produce retail, enabling real-time freshness tracking and allowing retailers to share valuable information—like sustainability certifications—directly with consumers. These innovations promise not only to streamline supply chains and reduce food waste but also to redefine competitive advantage. In a duopoly with freshness-sensitive demand, we explore whether an innovator—the unique adopter of supply transparency technology—gains a competitive advantage over a follower who adopts the technology when her competitor has adopted it. Our analysis reveals that the innovator can achieve higher profits, reduce food waste, enhance consumer satisfaction, and increase demand fulfillment. When consumers value transparency, the innovator's advantage grows, reshaping the market in her favor. Notably, the innovator curbs more food waste through a "market shrinkage" effect, where the competitor's market share contracts, easing supply-demand imbalances and mitigating the risk of excess inventory. Meanwhile, the follower's gains are limited, as the non-adopter passively benefits from the innovator's proactive choices. We also find that markets with fewer competitors are more inclined toward industry-wide adoption, suggesting that mergers in grocery retail could further amplify food waste reduction and consumer demand satisfaction. Finally, we extend our findings to cases involving freshness-based discounts and sequential technology adoption.

3 - Integrating AI, Remote Sensing, and Optimization for Urban Flood Resilience

Yuki Miura, New York University, Brooklyn, NY, United States

As climate-driven flood hazards worsen, both public and private institutions must make high-stakes investment decisions under deep uncertainty. We introduce an optimization framework for resilient flood risk management that integrates dynamic flood simulations with structural and socioeconomic exposure data to support financially sound resource allocation informed by stakeholders' inputs. A key contribution is the development of a novel global rooftop material dataset, derived from satellite imagery using machine learning, which serves as a scalable proxy for structural vulnerability. Together, the framework and dataset enable robust scenario-based planning: we simulate millions of plausible flood events and various strategies to identify interventions that remain effective across diverse futures. New York City has served as a testbed through collaborations with multiple agencies spanning administration, utility, transportation, and public safety. This work offers a new paradigm for operationalizing climate resilience at scale and guides actionable solutions.

4 - Electric Infrastructure Investment Under Uncertainty

Christoph Graf, New York University, New York, NY, United States, David Wozabal

Electric infrastructure investment decisions are increasingly challenged by deep uncertainty arising from climate change and extreme weather events, which impact both electricity demand and the supply from weather-dependent renewable energy sources. Enabling electric infrastructure, such as storage and transmission, can enhance the value of renewables by mitigating their variability across time and space. We present a stochastic electricity investment model for the U.S. that is highly granular in time and space. Our approach incorporates classical optimization methods alongside preliminary applications of deep learning techniques to explore investment decisions under a range of uncertain futures.

5 - Sustainable Investment Strategies with Real Asset Trades

Felipe Verastegui, Columbia University, New York, NY, United States, Agostino Capponi, Jay Sethuraman

What strategy should sustainable investors adopt to reduce emissions without inducing firms to excessively engage in the reallocation of industrial pollution? We study this question in a matching model in which entrepreneurs choose attributes for their businesses and then form stable matchings with investors. The presence of environmentally minded investors in the matching market affects the attribute choice creating regimes with pooling and separating equilibria, which depends on the strength of the sustainable preferences. When we account for real asset trades, there is a misalignment between value-based and impact-based preferences: entrepreneurs can achieve warm-glow in a separating equilibrium without reducing emissions, by efficiently reallocating their productive assets. Trades prove efficient for value-based investors, but limit the scope and attractiveness of impact investing strategies.

WA50

Building B Level 3 B310

Customer Behavior in Service Systems

Invited Session

MSOM: Service Operations

Chair: Jimmy Qin, University of Texas in Dallas, Richardson, United States

Co-Chair: Andrew Frazelle, The University of Texas at Dallas, Richardson, TX, United States

1 - Mitigating Service Incompletion via Sunk Costs: a Queueing Approach

Jimmy Qin, The University of Texas at Dallas, Richardson, TX, United States, Carri Chan, Jing Dong

Sunk-cost bias occurs when decisions are influenced by the time, energy, and money already invested, rather than considering the future costs necessary to achieve success. This phenomenon of "irrational behavior" is well-documented in decision-making studies and is generally recognized as a factor that can lead to suboptimal decisions. In this work, we investigate how sunk cost (and the behavioral bias associated with it) can be used as an operational lever to increase service completion rates in a congested service system. We run a controlled online experiment and find that the abandonment rate is significantly reduced for the group of participants who incur a larger sunk cost. To better capture the dynamics of service systems and their impact on customers' behavior, we study a queueing model with sunk cost and strategic customers, where customers experience a disutility of balking that is proportional to the sunk cost they incur. We characterize the equilibrium behavior of the customers, from which we further derive the optimal strategy for the service provider in terms of whether to provide real-time queue length information to customers as well as the optimal level of sunk cost to impose. Our results show that the sunk cost strategy is effective only when waiting information is provided and that using a non-zero sunk cost is optimal when the queueing system is moderately congested. Through a comprehensive numerical study, we demonstrate that implementing a non-zero sunk cost can substantially improve the throughput of the system.

2 - Bayesian Pricing for Impatient Customers with Unknown Valuation

Zhenghang Xu, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Philipp Afeche, Opher Baron, Dmitry Krass

We study a Bayesian dynamic pricing problem for a revenue-maximizing, capacity-constrained service provider that learns unknown customer valuations through sequential pricing. Customers arrive sequentially, accept or reject the prevailing price, and those who accept occupy the server; arrivals are lost if the server is busy. Our contributions include: (1) A novel recursive algorithm to compute optimal prices that balance demand learning and revenue maximization; (2) theoretical identification of an *adaptive explore-then-commit* structure: when uncertainty resolves sufficiently, the firm prices at the valuation's lower bound to guarantee sales; (3) a practical heuristic for delay-free settings, inspired by the Bayesian optimal policy, that numerically outperforms existing benchmarks and achieves the optimal regret order. By unifying Bayesian learning with pricing dynamics, we show how firms adaptively transition from exploration to exploitation. Results demonstrate how customer impatience and server constraints shape pricing strategies, offering actionable methods with provable guarantees for revenue management under uncertainty.

3 - Information Avoidance In Diagnostic Testing

Zack Zhu, University of British Columbia, Surrey, BC, Canada, Steven Shechter, Woonghee Huh

In formalizing information acquisition decisions as combined instrumental and hedonic utility maximization problems, we develop insights on the degree to which individuals seek vs. avoid diagnostic tests. By explicitly incorporating a psychological positive utility for a good test outcome and a negative utility for a bad test outcome, we provide a decision-analytic model that rigorously explains why for some people, "ignorance is bliss." We apply this model to the context of COVID-19, examining individuals' decisions related to testing. The model sheds light on willful ignorance during the pandemic, and offers policy-relevant insights into how public health interventions can help mitigate the societal costs due to information avoidance.

WA51

Building B Level 3 B311

Security and Resilience of Critical Systems: Advances in Optimization Under Strategic and Stochastic Uncertainty

Invited Session

OPT: Optimization Under Uncertainty

Chair: Mathieu Dahan, Georgia Institute of Technology, Atlanta, GA, United States

1 - Adaptive Sampling and Partition-Based Methods for Stochastic Shortest Path Interdiction Problems

Di Nguyen, University College Dublin, Dublin, Ireland

We investigate the computational performance of two solution approaches and various enhancements in a two-stage shortest-path network interdiction game. This game involves an interdicator and an evader. The evader seeks to traverse the network using a source-sink path that is shortest, while the interdicator seeks to maximize this cost by interdicting a subset of arcs given a budget. In the first stage, the interdicator determines an interdiction solution while only being able to anticipate the range of arc costs. In contrast, the evader can observe the exact arc cost realizations at the time of their operation, as well as the impact due to the interdicator's decision in the second stage. In this problem, we act as the interdicator and find an interdiction solution that maximizes the evader's expected shortest-path cost. Our research evaluates two main approaches to solving this problem, one leveraging the adaptive sampling framework and one utilizing partition-based methods. Several enhancements that exploit the problem's structure are explored. The computational results, along with their advantages and disadvantages, are discussed and compared against the existing benchmark from the work of Nguyen and Smith (2022).

2 - A Game-Theoretic and McdA Framework for Sustainable Arctic Development: Balancing Economic Interests, Security, and Indigenous Rights

Yiqi Zhao, University at Buffalo, SUNY, Buffalo, NY, United States, Jun Zhuang

The rapid economic development in the Arctic, driven by its natural resources and strategic significance, poses substantial challenges. This study integrates game theory and multi-criteria decision analysis (MCDA) to scientifically examine the interplay between economic growth and indigenous interests, enhancing the robustness and analytical rigor of decision-making. The research employs sequential and non-cooperative game theory to model strategic interactions among governments, corporations, and indigenous communities, while MCDA evaluates and prioritizes diverse criteria to assess policy impacts comprehensively. This combined approach aids policymakers in balancing economic objectives with the protection of indigenous rights, sustainability, and regional stability. Indigenous concerns regarding cultural heritage, traditional lifestyle disruption, resource overexploitation, and environmental risks are addressed through MCDA-driven, data-informed analysis. Recommendations include promoting indigenous participation in policy-making, ensuring fair distribution of economic benefits, enforcing strict environmental safeguards, and supporting long-term community development. The findings demonstrate strategies that respect indigenous sovereignty and promote sustainable development. Integrating game theory with MCDA equips policymakers with scientifically grounded tools for informed, collaborative decision-making, balancing economic goals with indigenous rights and environmental stewardship.

3 - Learn to Optimize: Self-Supervised Progressive Hedging and Adaptive Penalties for Multistage Stochastic Network Recovery Planning

Graham Hulsey, George Mason University, Fairfax, VA, United States, Shima Mohebbi

This work explores using self-supervised machine learning (ML) to solve large and computationally difficult multi-stage stochastic programs (MSSPs). Recent work integrating machine learning into existing algorithms to solve MSSPs has shown promise in reducing computational time with minimal reduction in solution quality. In the Learn-to-Optimize framework, ML models are used to either predict solutions or algorithm parameters directly, or to improve algorithm structure or model formulation via classification, prediction, or clustering models. Typically, these methods are supervised methods and so require generating training data, which can be computationally expensive. The self-supervised approach proposed in this work builds on the progressive hedging (PH) algorithm, a stochastic programming algorithm with well-known properties, and because it is self-supervised does not require training data. Moreover, the approach can seamlessly handle binary variables, which is a major bottleneck for classical solver-based algorithms. Computational experiments over random networks of different sizes demonstrate the self-supervised approach yields near-optimal solutions in substantially less time than PH with binary variables, though the self-supervised approach is sensitive to the choice of hyper-parameters. A real-world case study of Tampa, FL infrastructure networks is also presented.

4 - Wisdom: Wisconsin Intelligent System for the Detection of and Optimization Against Marine Threats

Vanessa Sawkmie, University of Wisconsin-Madison, Madison, WI, United States, Jeff Linderoth

The underwater domain presents significant challenges for surveillance and defense due to the inherent uncertainty in target behavior and detection outcomes. Effective detection depends on successful encounters between adversarial targets and search assets, each with distinct capabilities and operational constraints. Central to our approach is the concept of a search pattern, defined by a spatial region and time interval, to which different asset types can be assigned. To address uncertainty in both target motion and detection outcomes, we formulate a two-stage stochastic integer programming model that strategically allocates search assets across a decision theater over a finite planning horizon to maximize the expected number of targets detected. We extend this framework to a three-stage stochastic model by incorporating fixed bottom sensors that provide probabilistic observations of target movement, thereby informing more effective pattern assignments for the search assets.

WA52

Building B Level 3 B312

Computer-aided Design and Verification for First-order Optimization Algorithms

Invited Session

OPT: Computational Optimization and Software

Chair: Jisun Park, Princeton University, Princeton, NJ, 08540, United States

1 - Data-Driven Analysis of First-Order Methods via Distributionally Robust Optimization

Jisun Park, Princeton University, Princeton, NJ, United States, Vinit Ranjan, Bartolomeo Stellato

We consider the problem of analyzing the probabilistic performance first-order methods, when solving a distribution of parametric convex optimization problems. By combining performance estimation (PEP) and Wasserstein distributionally robust optimization (DRO), we formulate the analysis as a tractable semidefinite program. Our approach unifies worst-case and average-case analyses by incorporating data-driven information from the observed convergence of first-order methods on a limited number of problem instances. This yields probabilistic, data-driven performance guarantees—specifically, in terms of the expectation or conditional value-at-risk of the selected performance metric. Experiments on convex quadratic and smooth convex minimization show that our method significantly reduces the conservatism of classical worst-case bounds and narrows the gap between theoretical and empirical performance.

2 - Exact Verification of First-Order Methods via Mixed-Integer Linear Programming

Bartolomeo Stellato, Princeton University, Princeton, NJ, United States

We present exact mixed-integer linear programming formulations for verifying the performance of first-order methods for parametric quadratic optimization. We formulate the verification problem as a mixed-integer linear program where the objective is to maximize the infinity norm of the fixed-point residual after a given number of iterations. Our approach captures a wide range of gradient, projection, proximal iterations through affine or piecewise affine constraints. We derive tight polyhedral convex hull formulations of the constraints representing the algorithm iterations. To improve the scalability, we develop a custom bound tightening technique combining interval propagation, operator theory, and optimization-based bound tightening. Numerical examples, including linear and quadratic programs from network optimization, sparse coding, and optimal control, show that our method provides several orders of magnitude reductions in the worst-case fixed-point residuals, closely matching the true worst-case performance.

3 - Composing Optimized Stepsize Schedules for Gradient Descent

Benjamin Grimmer, Johns Hopkins University, Baltimore, MD, United States, Kevin Shu, Alex Wang

Recent works by Altschuler and Parrilo and the authors have shown that it is possible to accelerate the convergence of gradient descent on smooth convex functions, even without momentum, just by picking special stepsizes. In this talk, we will provide a general theory for composing stepsize schedules capturing all recent advances in this area and more. We propose three notions of "composable" stepsize schedules with elementary associated composition operations for combining them. From these operations, in addition to recovering recent works, we construct three highly optimized sequences of stepsize schedules. We first construct optimized stepsize schedules of every length generalizing the exponentially spaced silver stepsizes. We then construct highly optimized stepsizes schedules for minimizing final objective gap or gradient norm, improving on prior rates by constants and, more importantly, matching or beating the numerically computed minimax optimal schedules. We conjecture these schedules are in fact minimax (information theoretic) optimal. Several novel tertiary results follow from our theory including recovery of the recent dynamic gradient norm minimizing short stepsizes and extending them to objective gap minimization.

4 - Optimization Algorithm Design via Electric Circuits

Tetiana Parshakova, Amazon, New Haven, CT, United States, Stephen Boyd, Ernest Ryu, Jaewook Suh

We present a novel methodology for convex optimization algorithm design using ideas from electric RLC circuits. Given an optimization problem, the first stage of the methodology is to design an appropriate electric circuit whose continuous-time dynamics converge to the solution of the optimization problem at hand. Then, the second stage is an automated, computer-assisted discretization of the continuous-time dynamics, yielding a provably convergent discrete-time algorithm. Our methodology recovers many classical (distributed) optimization algorithms and enables users to quickly design and explore a wide range of new algorithms with convergence guarantees.

WA53

Building B Level 3 B313

Advances and New Applications of Choice Models

Invited Session

Revenue Management

Chair: Ruxian Wang, Johns Hopkins Carey Business School, Kensington, MD, 20895, United States

1 -

Ruxian Wang, Johns Hopkins Carey Business School, Kensington, MD, United States

2 - Sequential Choice Model with Representative Products: Modeling and Optimization

Yu SUN, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Daniel Zhuoyu Long, Ruxian Wang

Modeling decision behavior among multiple choice options has been an active research area for several decades. Typically, classic discrete choice models consider that customers observe all available products simultaneously and select the one with the highest utility (e.g., the widely used MNL and NL models). In this paper, we propose a novel sequential choice model with representative products to describe customers' purchase behaviors, in which the products are grouped in multiple nests, and each nest has a representative product that is viewed in the first stage and other products that are viewed in the second stage. Customers make their purchase decision sequentially following the product information disclosure process in two stages: in the first stage, customers only observe incomplete information about the product set (i.e., limited information of all representative products) and choose a nest with the most attractive representative or leave without purchasing; in the second stage, customers unlock full information about all products in the selected nest and purchase the product in this nest with the highest utility. We employ the widely used MNL choice model in each stage to formulate the sequential choice process. Under this sequential choice model, we first investigate the problem on the selection of representative products and product information display management. We then investigate the assortment optimization, nest planning and pricing problem, and establish the structural properties of optimal solutions. We finally present effective estimation techniques and comparative statics to demonstrate its application to real-world issues.

3 - Tax Relief OR Public Belief: CSR-Induced Consumer Segmentation with Tax Incentives

Xinyi ZHOU, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Lijian Lu, Ruxian Wang

We study a retail firm selling quality-differentiated products to both low- and high-expertise consumers and analyze how its level of Corporate Social Responsibility (CSR) investment influences consumer preferences and purchase behavior. CSR efforts enhance low-expertise consumers' confidence in the firm but have no impact on high-expertise consumers, leading to market segmentation. When high quality is relatively inexpensive, high-expertise consumers purchase only the high-quality product, while low-expertise consumers are split across product types. CSR and pricing strategy together can drive market-level quality upgrading by shifting demand from low- to high-quality products. We further examine how government tax deduction policies interact with CSR investment and affect the welfare of the firm, consumers, and society. We find that the benefit of charity-related tax deductions depends critically on who holds decision power and how the charitable tax relief scheme is designed.

4 - Product Information Display Optimization with Quality Uncertainty and Returns

Yang Lu, University of Science and Technology of China, Hefei, China, People's Republic of, Ruxian Wang, Xiaoming Yan, Yugang Yu

Display quality refers to product information such as the images and textual descriptions of product quality presented to consumers shopping online, and can directly influence market demand during the sales phase and product returns during the after-sales phase. In this paper, we investigate the product display optimization problems on display quality, pricing, and assortment planning, taking into account the uncertainty of product quality and consumer returns. We employ the widely used multinomial logit (MNL) model to investigate the impact of display quality on consumer choices. In particular, we incorporate display quality into the utility function regarding consumers' purchase and return behavior, which is largely affected by display quality. We then conduct a joint optimization analysis on assortment planning, display quality and pricing, and characterize the optimal solution and propose an efficient method that involves assessing whether the conservative retention rate of a product is positive. Moreover, we investigate the estimation problem and conduct a comprehensive empirical validation with real data. Our results reveal that model misspecification and optimization on display quality could result in substantial losses. Because of its high impact on both market demand and subsequent returns, display quality alters the structure of the optimal pricing and assortment decisions. In addition, we find that optimizing display quality creates another opportunity to achieve a win-win outcome for both the retailer and consumers, which could serve as a reference for market regulation.

WA54

Building B Level 3 B316

Smart Charging and Clean Mobility Systems

Contributed Session

Contributed

Chair: Sandjai Bhulai, Vrije Universiteit Amsterdam, De Boelelaan 1111, Amsterdam, 1081 HV, Netherlands

1 - Energy-Aware Resource Allocation for Industry 4.0 Using Evolutionary Game Theory**Ka Shuen Ho, Purdue University, West Lafayette, IN, United States, Nga Hin Ben Fong**

In the context of Industry 4.0, smart factories operate in highly dynamic environments with fluctuating workloads and energy constraints. This requires real-time, adaptive resource allocation strategies. At the same time, energy-aware operations have become essential for minimizing environmental footprint, reducing costs, and meeting strict energy regulations. Traditional centralized approaches often lack the scalability and responsiveness needed under such conditions, prompting a shift toward decentralized, autonomous decision-making systems. To address these challenges, this study proposes an evolutionary game theory (EGT) framework for energy-aware resource allocation in smart factories to achieve greater efficiency, flexibility, and sustainability. In this framework, each machine is modelled as an autonomous agent that iteratively evolves its task selection and operation strategies through game-theoretic dynamics. Payoffs incorporate multiple factors including energy consumption, task priorities, and resource requirements. Through local interactions and strategic adaptation, the system converges to equilibria at evolutionarily stable strategies (ESS) that balance competing objectives such as throughput and energy efficiency. The research demonstrates the potential of EGT-based approach for supporting sustainability and operational performance in smart manufacturing systems.

2 - Coordinated Load Shifting Through Storage-Enabled Demand Response in Smart Grids**Ahmed Ali, University of Miami, Miami, FL, United States, Murat Erkoç**

With growing emphasis on demand-side flexibility and the increasing availability of behind-the-meter storage, effective demand management through dynamic pricing has become essential for ensuring the reliability and efficiency of modern electricity systems. This study investigates the impact of consumer-level electricity storage on improving load shifting strategies within a demand-side management framework. A Stackelberg game is established between a single electricity provider and a large number of heterogeneous consumers. In this model, the provider implements time-varying price discounts to encourage consumers to shift their usage from peak to off-peak periods. Consumers, in turn, independently optimize their electricity consumption, considering their individual demand preferences and inconvenience costs. Two control paradigms are analyzed: provider-controlled storage and consumer-controlled storage. Through analytical characterization and simulation outcomes, the study reveals how the integration and management of storage influence load shifting behavior, peak to average demand ratio reduction, and overall economic performance. The findings suggest that consumer diversity regarding inconvenience levels significantly affects the equilibrium outcomes. Additionally, the incorporation of storage enhances the efficacy of price-based load shifting strategies, with provider control enabling more coordinated peak management, while consumer control provides greater flexibility and welfare benefits. These insights contribute to the formulation of effective storage deployment strategies and dynamic pricing policies in smart grid systems.

3 - Scheduling Charging of Electric Vehicles at Airport Rental Car Centers**Devon Sigler, National Renewable Energy Laboratory, Longmont, CO, United States, Gustavo Campos, Juliette Ugirumurera, Yanbo Ge, Monte Lunacek, Caleb Phillips, Roberto Vercellino, Margaret Mann, Yi He**

Airports and rental car companies are currently grappling with the infrastructure and operational challenges of fleet electrification. Full electrification of an airport rental car fleet requires charging to be integrating into their operational model. Charging vehicles using DC fast charging allows rental car companies to minimally alter their operational paradigm. However, plugging vehicles into DC fast charging infrastructure immediately upon arrival will create new electrical loads in the tens of Megawatts at major airports, which the current power grid infrastructure is not equipped to support. In this work we investigate the potential of smart charge scheduling to the reduce peak loads at airport rental car centers from vehicle charging. Towards this end we present two scalable smart charge scheduling algorithms and demonstrate their impact through simulation of real-world scale vehicle fleets.

4 - Optimizing Power Grid Topologies with Reinforcement Learning: a Survey of Methods and Challenges**Sandjai Bhulai, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, Erica van der Sar, Alessandro Zocca**

Power grid operation is becoming increasingly challenging due to the growing integration of renewable energy sources and the demand for more adaptive control strategies. Reinforcement Learning (RL) offers a promising solution for Power Network Control (PNC), with its ability to support effective decision-making under dynamic and uncertain conditions. The Learning to Run a Power Network (L2RPN) competitions have significantly contributed to the field by offering standardized benchmarks and well-defined problem settings, thereby accelerating research progress. This survey presents a structured and comprehensive review of RL applications in power grid topology optimization. It categorizes current methodologies, examines critical algorithmic design decisions, and identifies key research gaps. Furthermore, it includes a comparative numerical analysis of frequently used RL approaches, evaluating their practical performance and limitations. By synthesizing the current state of the field and outlining unresolved challenges, this survey provides a valuable reference for guiding future developments in RL-based power grid optimization.

WA55

Building B Level 3 B314

Online Decision Making for RM Problems

Invited Session

Revenue Management

Chair: Zizhuo Wang, CUHK Shenzhen, Shenzhen, N/A

1 - Wait-Less Offline Tuning and Re-Solving for Online Decision Making**Jingruo Sun, Stanford University, Stanford, CA, United States, Wenzhi Gao, Ellen Vitercik, Yinyu Ye**

Online linear programming (OLP) has found broad applications in revenue management and resource allocation. State-of-the-art OLP algorithms achieve low regret by repeatedly solving linear programming (LP) subproblems that incorporate updated resource information. However, LP-based methods are computationally expensive and often inefficient for large-scale applications. In contrast, recent first-order OLP algorithms are more computationally efficient but typically suffer from worse regret guarantees. To address these shortcomings, we propose a new algorithm that combines the strengths of LP-based and first-order OLP methods. The algorithm re-solves the LP subproblems periodically at a predefined frequency f and uses the latest dual prices to guide online decision-making. In addition, a first-order method runs in parallel during each interval between LP re-solves, smoothing resource consumption. Our algorithm achieves $O(\log(T/f) + \sqrt{\log T})$ regret, delivering a "wait-less" online decision-making process that balances the computational efficiency of first-order methods and the superior regret guarantee of LP-based methods.

2 - Beyond Regularity: Simple Versus Optimal Mechanisms, Revisited**Yiding Feng, Hong Kong University of Science and Technology, Hong Kong, Hong Kong, Yaonan Jin**

A large proportion of the Bayesian mechanism design literature is restricted to the family of regular distributions [Mye81] or the family of monotone hazard rate (MHR) distributions [BMP63], which has overshadowed this rich and well-developed theory. We (re-)introduce two generalized families: quasi-regular distributions and quasi-MHR distributions. Likewise, the parameterized families of λ -regular (a.k.a. α -strongly regular) distributions [CR14,SS19], which smoothly interpolate regular distributions and MHR distributions, generalize to λ -quasi-regular distributions.

The significance of our new families is manifold. Firstly, their defining conditions are immediate “economic” relaxations of the original defining conditions (e.g., regularity as monotonicity of the virtual value functions), capturing key economic intuitions. Secondly, they satisfy natural mathematical properties (about order statistics) failed for the original families, thus technically more tractable. Thirdly, numerous results (by [BK96, HR09, CD15, DRY15, HR14, AHNPY19, JLTx20, JLQTX19, FLR19, GHZ19, JX23, LM24] etc) known merely for the original families now can extend to our new families. Many of these extensions incur no quantitative loss, or even improve the state of the art for the original families. Finally, beyond the third point, our new families guide us to entirely new perspectives and thus entirely unknown

results. For example, regarding revenue maximization for symmetric versus asymmetric regular buyers, we acquire 0.5- versus 0.1908-approximations for the (less-than-)one-sample prophet inequalities, respectively. To the best of our knowledge, such results are blank in the literature, despite their widely-studied welfare maximization counterparts [CDFS19, RWW20, CCES20, CDFS21, CCES24].

3 - Instance-Dependent Sample Complexity for Bilinear Saddle-Point Optimization with Noisy Feedback: An Lp-Based Approach

Jiashuo Jiang, HKUST, Hong Kong, Hong Kong

Abstract: In this work, we study the sample complexity of finding a Nash Equilibrium (NE) of two-player zero-sum matrix games with noisy feedback. We propose a novel algorithm based on resolving linear programs (LP), achieving an ϵ -approximate NE with an instance-dependent sample complexity for general m_1 times m_2 game matrices, while being robust by enjoying a worst-case sample complexity at the same time. To our knowledge, this result provides the first instance-dependent sample complexity for general-dimension matrix games under noisy feedback and non-unique NE. Our algorithm is inspired by recent advances in online resource allocation and includes two key components: first, identifying a support set of the NE, and second, determining the unique NE within this support. Both steps rely on a careful analysis of the performance of solving the LP constructed by noisy samples. This is a joint work with Mengxiao Zhang

4 - Llm Serving with Variable Prefill & Decode Lengths

Meixuan Wang, Tsinghua University, Beijing, China, People's Republic of, Zijie Zhou

We study the problem of serving LLM (Large Language Model) requests where each request has heterogeneous prefill and decode lengths. In LLM serving, the prefill length corresponds to the input prompt length, which determines the initial memory usage in the KV cache. The decode length refers to the number of output tokens generated sequentially, with each additional token increasing the KV cache memory usage by one unit. Given a set of n requests, our goal is to schedule and process them to minimize the total completion time. We show that this problem is NP-hard due to the interplay of batching, placement constraints, precedence relationships, and linearly increasing memory usage. We then analyze commonly used scheduling strategies in practice, such as First-Come-First-Serve (FCFS) and Shortest-First (SF), and prove that their competitive ratios scale up with the memory limit—a significant drawback in real-world settings where memory is large. To address this, we propose a novel algorithm based on a new selection metric that efficiently forms batches over time. We prove that this algorithm achieves a constant competitive ratio of 48. Finally, we develop heuristic algorithms inspired by this approach and demonstrate through simulations that they outperform standard baselines.

5 - Online Linear Programming with Batching

Haoran Xu, Stanford University, Palo Alto, CA, United States

We study Online Linear Programming (OLP) with batching. The planning horizon is cut into K batches, and the decisions on orders arriving within a batch are permitted to be delayed to the end of their associated batch. Compared with OLP without batching, the ability to delay decisions improves operational performance, as measured by regret. We are interested in (1) What is a lower bound on the regret as a function of K ? (2) What algorithms achieve the regret lower bound? We analyze these questions when the distribution of the reward has a continuous support. With a single resource and a known total number of orders, we show a $\Omega(\log K)$ regret lower bound and propose an algorithm achieving this lower bound. We also propose algorithms with regret $O(\log K)$ for the setting with multiple types of resources and the setting with a Poisson order arrival process. All these regret bounds are independent of the length of the planning horizon, and all the proposed algorithms delay decisions on orders arriving in only the first and the last batch. We also take customer impatience into account and establish an appropriate batch size.

6 - Infrequent Resolving Algorithm for Online Linear Programming

Zizhuo Wang, CUHK Shenzhen, Shenzhen, China, People's Republic of

Online linear programming (OLP) has gained significant attention from both researchers and practitioners due to its extensive applications, such as online auction, network revenue management, order fulfillment and advertising. Existing OLP algorithms fall into two categories: LP-based algorithms and LP-free algorithms. The former one typically guarantees better performance, even offering a constant regret, but requires solving a large number of LPs, which could be computationally expensive. In contrast, LP-free algorithm only requires first-order computations but induces a worse performance, lacking a constant regret bound. We bridge-the-gap between these two extremes by proposing well-performing algorithm, that solves LPs at a few selected time points and conducts first-order computations at other time points. Specifically, for the case where the inputs are drawn from an unknown finite-support distribution, the proposed algorithm achieves a constant regret (even for the hard “degenerate” case) while solving LPs only $O(\log \log T)$ times over the time horizon T . Moreover, when we are allowed to solve LPs only M times, we design the corresponding schedule such that the proposed algorithm can guarantee a nearly $O(T^{\frac{1}{2}(M-1)})$ regret. Our work highlights the value of resolving both at the beginning and the end of the selling horizon, and provides a novel framework to prove the performance guarantee of the proposed policy under different infrequent resolving schedules. Furthermore, when the arrival probabilities are known at the beginning, our algorithm can guarantee a constant regret by solving LPs $O(\log \log T)$ times, and a nearly $O(T^{\frac{1}{2}M})$ regret by solving LPs only M times.

WA56

Building B Level 3 B315

AI, Information, and Algorithmic Market Design for Platforms

Invited Session

Auctions and Market Design

Chair: Rad Niazadeh, Chicago Booth School of Business, Chicago, IL, United States

Co-Chair: Jiaqi Lu, the Chinese University of Hong Kong (Shenzhen), Shenzhen, N/A

1 - Information Design and Pricing In Lead-Selling Platforms with Supply Competition

Jiahua Wu, Imperial College London, London, United Kingdom, Yanwei Sun, Niloofar Zamani Foroushani, Zhe Liu

We analyze a type of platform common in the service sector, where customers can freely post tasks, and service providers pay a fee to purchase job leads and unlock the ability to communicate with these customers. We adopt a Bayesian persuasion framework to explore how the platform can leverage its informational advantage about the provider base and pricing to strategically influence the decisions of service providers, aiming to maximize its revenue. We begin the analysis by assuming that prices are exogenously given and characterize the structure of the optimal signaling mechanism. Specifically, the platform recommends that all service providers participate when the pool size is below a certain threshold. When the pool size exceeds this threshold, the platform randomizes its recommendations between two consecutive numbers, independent of the pool size. This behavior is driven by a unique feature of our model, where the beliefs of the platform and service providers are non-identical: a provider's knowledge of their own participation inflates their belief about the total number of service providers. Next, we endogenize the platform's pricing decision and consider the policy space where the signaling mechanism and (possibly randomized) state-dependent pricing are jointly optimized. Notably, within this space, we find that the optimal signaling mechanism with fixed pricing is optimal. This result implies the superiority of the optimal signaling mechanism over the commonly adopted contingent pricing policy. We show that the performance discrepancy between these two policies is particularly pronounced when supply competition is intense.

2 - Optimizing Pilot Mentorship Through Algorithmic Matching: Evidence from a Randomized Controlled Trial

Meixi Wu, National University of Singapore, Singapore, Singapore, Jussi Keppo, Hong Ming Tan

This study evaluates the effectiveness of mentoring workshops and algorithmic matching in enhancing pilot performance and mentoring quality through a randomized controlled trial (RCT). Drawing on historical data, we first identify key features correlated with improved performance, then design and implement an integer programming-based algorithm that optimizes mentor-mentee pairings using a weighted combination of these features while satisfying operational constraints. Compared to conventional matching, algorithmic pairing leads to higher post-workshop competency scores, as shown by an analysis of covariance (ANCOVA) that controls for pre-workshop performance and a Difference-in-Differences (DID) model with mentee fixed effects to account for time-invariant individual heterogeneity. Survey evidence further shows that algorithmic matching enhances mentees' perceived mentoring quality (as measured by the LMX framework) and strengthens perceptions of legitimate power. These findings demonstrate that data-driven matching can enhance mentorship effectiveness in aviation training.

3 - AI Agents Can Enable Superior Market Designs

Gili Rusak, Harvard University, Cambridge, MA, United States, Benjamin Manning, John Horton

Many theoretically appealing market designs are under-utilized because they demand preference data that humans find costly to provide. This paper demonstrates how large language models (LLMs) can effectively elicit such data from natural language descriptions. In our experiment, human subjects provide free-text descriptions of their tastes over potential roles they could be assigned. An LLM converts these descriptions into cardinal utilities that capture participants' preferences. We use these utilities and participants' stated preferences to facilitate three allocation mechanisms---random serial dictatorship, Hylland-Zeckhauser, and a conventional job application type game. A follow-up experiment confirms that participants themselves prefer LLM-generated matches over simpler alternatives outside their top choices---exactly the choices market participants face under high congestion. These findings suggest that LLM-proxied preference elicitation can enable superior market designs where they would otherwise be impractical to implement.

4 - Strategic Supply-Demand Mismatch through Lead Time and Pricing Autonomy: Evidence from a Car-Sharing Platform

Asif Nazrul, University of Cambridge, Cambridge, United Kingdom, Stylianos Kavadias

This paper examines supply-demand mismatch on hybrid peer-to-peer (P2P) car-sharing platforms, where decentralized hosts make strategic decisions about asset availability. Using booking-level data from one of largest online marketplace and exploiting a quasi-experimental policy shift in February 2023 that granted pricing autonomy to hosts, this study investigates how self-pricing affects listing behavior and utilization. A Triple Difference-in-Differences (DDD) framework identifies heterogeneous treatment effects moderated by host experience.

The results show that experienced hosts with self-pricing privileges strategically delay their listings, reducing lead times by an additional 0.92 days compared to other host groups. This temporal shift contributes to a 6.9 percentage point decline in utilization, reflecting lower booked hours relative to availability. Inexperienced self-pricing hosts also experience diminished performance, likely due to suboptimal pricing decisions. Together, these findings suggest that pricing autonomy encourages strategic but inefficient supply behavior, particularly among experienced providers.

This study contributes to the literature on marketplace design and decentralized operations by offering causal evidence on the behavioral consequences of pricing flexibility. It highlights how autonomy can exacerbate service availability gaps, particularly in advance-booking settings where supply timing is a critical decision that is overlooked in current literature. The findings underscore the need for experience-sensitive interventions, such as feedback-based incentives or dynamic listing nudges, to maintain platform efficiency while preserving host flexibility. Future work can extend this analysis using unsupervised clustering to uncover latent experience profiles and improve behavioral targeting in two-sided digital markets.

5 - Optimizing Mixed Workforce in Delivery Platforms

Xinyuan Wu, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China, People's Republic of, Yunpeng Li, Costas Courcoubetis

Delivery platforms rely on full-time couriers in low-demand periods, but during peak times, they supplement their workforce with crowdsourcing couriers. We study a platform operating both courier types during peak periods. While full-time couriers are centrally dispatched, crowdsourcers strategically choose the number of orders to deliver simultaneously to maximize their average rewards, where the job waiting time is endogenously determined by their collective behavior. The crowdsourcer interaction corresponds to a mean-field game where agents compete for sequentially consumable resources. We show that their behavior forms a submodular game and the equilibrium is characterized by a convex problem. To optimize platform profit, we formulate a Stackelberg game where the platform acts first to decide on the demand allocated to full-time and crowdsourcers, then crowdsourcers respond by reaching equilibrium. To avoid directly solving this bilevel problem, we analytically derive and compare the marginal profit of serving additional demand via full-time couriers and crowdsourcers, proposing an efficient heuristic for optimal demand allocation. On a longer time scale, crowdsourcers observe the demand assigned to them and make their participation decisions. Crowdsourcers require both sufficient allocated demand and adequate per-job payment to engage. The platform can strategically adjust payment to expand participation when profitable. We extend our heuristic to jointly optimize demand allocation and crowdsourcer payment. Our results provide insights into the strategic behavior of crowdsourcers, offering scalable solutions for demand allocation and balancing operational efficiency with crowdsourcing incentives.

WA58

Building B Level 3 B317

Healthcare and Social Impact

Contributed Session

Contributed

Chair: Zhongxin Hu, Eindhoven University of Technology, Eindhoven, N/A, Netherlands

1 - Context-Driven Sustainability Metrics for Smes: Smart Decision Frameworks in Resource-Constrained Manufacturing

Akshay Gajanan Khanzode, Indian Institute of Management Visakhapatnam, Visakhapatnam, India

In emerging markets, small and medium-sized manufacturing enterprises (SMEs) are critical nodes in global supply chains, yet they face acute challenges in adopting sustainability practices due to limited resources, technical expertise, and data infrastructure. Traditional ESG frameworks (e.g., GRI, SASB) often overwhelm SMEs with their complexity and misalignment with ground-level realities.

This study proposes a decision-analytic approach tailored to resource-constrained firms by integrating Context-Based Sustainability (CBS) and the MultiCapital Scorecard (MCS)—frameworks that assess environmental and social performance against ecological thresholds and stakeholder expectations. Through detailed case studies from Indian SMEs we demonstrate how SMEs translate abstract sustainability goals into actionable, localized metrics that improved both performance and resilience.

The presentation highlights a five-step playbook to help SME leaders identify material issues, apply open-source metrics, and embed sustainability into operational decision-making. This pragmatic framework bridges the gap between aspirational sustainability and data-informed strategy, offering a scalable model for other emerging market firms.

2 - Carbon Trading and Offsetting with Eco-Conscious Consumer

Lingchunzi Li, Huazhong Agricultural University, Wuhan, China, People's Republic of

This study evaluates the effectiveness of carbon trading and offsetting mechanisms in a market with ecologically conscious consumers, which is vital for informing policy-making and enhancing social welfare. We analyze the interaction between these mechanisms and the impact of consumer environmental awareness on firms' low-carbon investments and pricing strategies in a duopoly market. Using a game-theory model that includes two producers with different emission reduction costs and eco-conscious consumers, we find that both carbon trading and offsetting encourage producers to invest more in emission reduction, with carbon trading leading to higher product prices. However, the integration of carbon offsetting can mitigate this price increase. Our findings challenge the conventional view by showing that in markets with high initial carbon emissions, increasing consumer environmental awareness may actually decrease total social welfare, including both economic and environmental benefits. Conversely, in markets with low initial carbon emissions, a combination of carbon trading and heightened consumer awareness can significantly enhance social welfare. This research addresses critical gaps in understanding the combined effects of carbon trading and offsetting, providing a comprehensive framework for more effective policy design in high energy-consuming industries.

3 - Identifying Trends in Logistics Facilities in Chicago and Their Impact on the Surrounding Communities

Otgondulam Boldbaatar, University of Illinois at Chicago, Chicago, IL, United States, Carla Tejada

The increasing demands of e-commerce and consumer expectations for rapid delivery have led to the growing importance of proximity logistics, prompting companies to position warehouses closer to city centers. While this strategic placement improves logistical efficiency by reducing transportation costs and delivery times, it also brings challenges such as proximity to vulnerable populations, increased exposure to noise and emissions, heightened health risks like asthma, and greater traffic congestion near logistics facilities.

This study examines e-commerce logistics facilities in the Chicago MSA, spanning Illinois, Indiana, and Wisconsin, with data from MWPVL International. To assess societal and environmental impacts, this study follows the methodology developed by Tejada and Conway (2024), identifying vulnerable populations near logistics facilities using variables such as minority status, poverty, unemployment, as well as added factors like car and housing access. The environmental analysis also follows this framework, using three transportation-related indicators: particulate matter 2.5 (PM_{2.5}), diesel particulate matter (DSL_M), and ozone (O₃) with data sourced from the EPA's Office of Environmental Justice. Traffic data was additionally obtained from the Chicago Open Data, which provides historical records for vehicle and truck volumes in the Chicago area.

Preliminary results indicate that in urban areas, e-commerce warehouses and logistics facilities are often located near vulnerable populations. Areas permitting such developments tend to show greater vulnerability than those that do not. Additionally, emissions are higher in low-income, minority communities and decrease in predominantly white, wealthier areas.

4 - Heart for Hearts: Pricing and Donation Strategies in the BoP Market with Prosocial Consumers

Xinyue Cai, Tsinghua University, Beijing, China, People's Republic of, Yongbo Xiao, Tony Cui, Chen Hu

With the growing emphasis on common prosperity and poverty alleviation, the affordability and availability of products in the underserved "Bottom of the Pyramid (BoP)" markets have emerged as a critical public concern. This research explores how firms strategically make optimal decisions by addressing this challenge, with socially responsible initiatives like price markdowns and product donations in the BoP market. In this paper, we examine a firm operating across two markets: the economically developed mainstream market and the less developed BoP market. By implementing markdowns and donations, the firm benefits from the increased willingness-to-pay of consumers in the mainstream market due to their prosocial nature. We analyze how mainstream market consumers' prosociality affects the firm's pricing decisions in both markets and the donation strategy in the BoP market. We uncover an intricate interplay between pricing and donation strategies, where donation can either complement or substitute for price markdown. With moderate consumer prosociality, the firm even charges a lower price together with donations. Surprisingly, our findings reveal that mainstream consumers' prosociality may inadvertently hurt consumers in both markets. We also extend the base model to capture BoP consumers' strategic behavior and examine how it may influence the firm's decision-making and stakeholders' payoffs.

5 - Global Agricultural Trade Under Seasonal and Policy Asymmetries: a Perspective from India's Pulse Imports

Shivam Sharma, Department of Management Sciences, IIT Kanpur, Kanpur, India, Sobhan Arisian, Vipin Balakrishnan, Raghu SENGUPTA

India's rising demand for pulses, driven by a projected population increase of 300 million between 2030 and 2050, shifting dietary preferences toward plant-based protein, and a predominantly vegetarian population, continues to outpace domestic production. Despite being the world's largest producer and consumer of pulses, India remains heavily reliant on imports—particularly lentils from countries such as Australia and Canada. This dependence is exacerbated by seasonal variability in global pulse production, inconsistent domestic yields, and reactive, often asymmetric, trade policy adjustments. These dynamics create significant uncertainty for exporters and introduce inefficiencies into the global agricultural supply chain.

This study examines the interaction between production seasonality and trade policy design in shaping import volumes and domestic market behavior, using India's pulse sector as a focal lens. We develop a sequential game-theoretic model capturing the strategic decisions of profit-maximizing importers and domestic producers under symmetric and asymmetric trade policy scenarios. In particular, we consider key policy mechanisms including simple tariff rate systems, tariff-rate quota (TRQ) systems, and their implementation via licensing and first come-first-serve (FCFS) allocation. These mechanisms introduce differentiated treatment across exporters, shaped by geopolitical factors, bilateral relationships, and domestic supply needs.

The analysis highlights how seasonality in global harvesting schedules and policy asymmetries jointly influence sourcing, production timing, and market outcomes. The framework provides actionable insights for policymakers seeking to balance domestic supply objectives with stable import flows, and for market participants navigating seasonal production asymmetries and evolving policy environments in global agricultural trade.

6 - Socially Responsible Operations for Medical Devices

Zhongxin Hu, Eindhoven University of Technology, Eindhoven, Netherlands, Yeqing Zhou, Christina Imdahl, Zumbul Atan

Access to essential medical devices (MDs) remains a persistent challenge for underserved communities. The World Health Organization estimates that up to 70% of donated equipment in low-resource countries is non-functional or underutilized due to high acquisition costs, inadequate infrastructure, and maintenance barriers. In response to these challenges, we propose a business model in which manufacturers refurbish and rent MDs to resource-limited communities using a pay-per-use (PPU) scheme. Each device has a finite lifetime, and manufacturers must decide when to refurbish in the developed community and how to price usage post-refurbishment in the underserved community. We formulate this decision-making problem as a Markov Decision Process (MDP) under dynamic refurbishment costs. We find structural properties of the optimal refurbishment timing and pay-per-use price. Numerical experiments show that optimal decisions are highly sensitive to the refurbishment cost. The results highlight that strategic timing and pricing in MDs transfer can extend asset utility, reduce waste, and foster a sustainable ecosystem that benefits both communities. Our work offers actionable insights for healthcare authorities and manufacturers seeking scalable solutions to address medical device shortages in resource-limited settings.

WA59

Building B Level 3 B318

Machine Learning in Healthcare

Contributed Session

Contributed

Chair: Ilker Guven, University of Louisville, Louisville, KY, United States

1 - Data Envelopment Analysis to Measure the Hospitals Efficiency: a Brazilian Case Study

REINALDO GARCIA, University of Brasilia, Brasilia, Brazil, Raphael Akira Hirata Iwamoto, Sofia André Pereira, Carlos Augusto de Mello Ferraz, Leonardo Gomes Ferreira, Luis Eduardo Weaver

The Brazilian universal public health system (SUS – *Sistema Único de Saúde*, Unified Healthcare System) has not only a budget shortcut but also a continuous growth in demand for healthcare services. It is essential then to identify ways to improve the program's performance without requiring additional investments, prioritizing a more efficient management of available resources. Therefore, the Brazilian Federal Accounting Office (TCU – *Tribunal de Contas da União*) has developed tools to promote efficiency and equity gains in hospital care. This study evaluates the implementation of these tools in Brazilian public hospitals applying the Data Envelopment Analysis (DEA) approach. The data set includes parameters of the number of beds and hours worked by healthcare professionals, geographical location, size, and the management level, among other parameters. The data was collected for about 2,941 hospitals from January 2019 to March 2024. A clustering technique based on Euclidean distance was applied to group hospitals with similar procedures, enabling comparisons among healthcare facilities. The preliminary results show that large hospitals having more than 300 ward beds are more efficient ones when compared with hospitals having less than 50 ward beds.

2 - Multi-Task Learning with Hybrid Feature Selection for Joint Prediction of Nafld and Fibrosis Staging Using Nhanes Data

Elmira Ahmadinedamani, Oklahoma State University, Stillwater, OK, United States, Tieming Liu

Nonalcoholic Fatty Liver Disease (NAFLD) is a major public health concern that can progress to advanced fibrosis if left undetected. We propose a novel machine learning framework that jointly predicts NAFLD status (binary classification) and fibrosis stage (ordinal regression) using data from the National Health and Nutrition Examination Survey (NHANES). Our model employs a Multi-Task Learning (MTL) architecture with CORAL loss to simultaneously optimize both tasks. To enhance interpretability and generalizability, we develop a hybrid feature selection pipeline that combines statistical, information-theoretic, and model-based approaches. This model-agnostic pipeline identifies a compact, clinically meaningful set of predictors by filtering redundant and non-informative features. Across 5-fold cross-validation, the model achieved an average AUC-ROC of 0.944, F1 score of 0.933, and Recall of 0.97 for NAFLD classification. For fibrosis staging, the model obtained a mean squared error of 0.283 and consistent class-wise performance. These results demonstrate strong predictive performance while maintaining interpretability, supporting early screening and risk stratification of liver disease in real-world, imbalanced population datasets.

3 - Using Mathematical Modeling to Understand Spatial Dynamics of Cortical Oxygen Consumption from fNIRS Data

Jun Hu, University of Michigan, Ann Arbor, MI, United States

Functional near-infrared spectroscopy (fNIRS) is a low-cost, portable, and user-friendly neuroimaging tool for detecting hemodynamic responses linked to brain activation via neurovascular coupling. It can be a great tool for building a digital twin of the dynamic brain functions and energy consumption in real-world settings. We introduce a novel methodology based on the laws of conservation of mass to estimate spatially and temporally dynamic oxygen consumption and distribution patterns of the cortical areas of the brain. We first use deep learning to automate motion artifact removal from the resting-state data. We then apply stochastic differential equations and Ito's Lemma to derive geometry-dependent correction factors to estimate concentrations of oxygenated and deoxygenated hemoglobin. We then use the temporal dynamics of the fNIRS data and law of conservation of mass to obtain point estimates of oxygen consumption and flow, and then finally use tessellation techniques and statistical modelling to obtain the spatial dynamics of this data from the point estimates. This data will finally be used to build a deep learning neural networks model to predict hemodynamic activity of the cortical brain. We believe that our work will produce a powerful diagnostic tool for identifying abnormalities in hemodynamic functions. Our early findings show that the estimated oxygen consumption and flow is synchronous with background physiological activity (e.g., heart rate, Meyer waves, etc.), supporting the physiological validity of our model. These results suggest the method successfully captures underlying blood flow dynamics driven by cardiac cycles.

4 - Survival Models for Children with Severe Acute Malnutrition: A Nigerian Case Study

Paul Rubin, Michigan State University, East Lansing, MI, United States, Ricardo Casonatto, Reinaldo C. Garcia, Jennifer Ostrowski, Brianna Eales, Tiago Teixeira, BRIAN CHAPLIN

Adequate nutrition is a critical determinant of health and development being vital for the survival of children under 5 years of age. Globally, an estimated 49 million (7.3%) children had acute malnutrition, with 16 million children being victims of severe acute malnutrition (SAM), accounting for estimated three million child deaths. In particular, it is estimated that 4.4 million children younger than 59 months suffered from acute malnutrition from May 2023 to April 2024 in northeast Nigeria. The SAM cases show failure to receive adequate nutrition due to episodes of diarrhea and other acute illnesses. It is then important to increase the numbers of Community Management of Acute Malnutrition (CMAM) centers in countries having a high prevalence of SAM. The CMAM centers allows to reduce the number of children who develop SAM and death complications. In order to aid the CMAN centers, this work implements statistical methods for survival data analysis including the models of Cox Proportional Hazard (CPH) and the Support Survival Vector Machine (SSVM). This work applies the CPH and the SSVM models to a sample of not only more than 5,000 Nigerian children but also with more than 1,400 initial parameters. These children take part in the Taimaka Program. The models results look very promising having an accuracy rate of over 0.83 and the implemented approaches do not have more than 27 final parameters showing their simplicity and robustness. This project is part of the Informs Pro-Bono Analytics initiative.

5 - A Study of Accessibility to Community Pharmacy Using Mobile Data and Machine Learning Models

Ilker Guven, University of Louisville, Louisville, KY, United States, Lihui Bai, Sumei Zhang

The COVID-19 pandemic exposed significant healthcare disparities, particularly in access to care. On the other hand, pharmacies as accessible healthcare points are known to be visited twice as often as primary care offices. This study aims to identify determinants in access to community pharmacy, thus helping to reduce health disparity. We employ machine learning models (e.g., linear regression, random forest, support vector machines, and multilayer perceptron) to predict annual visits and distance traveled to community pharmacies. Predictors include demographic, socioeconomic and geographic factors, as well as built environment and business features (e.g., distance to highway exits, % of service establishments, total number of establishments, crime rate). Diversity subsampling is used to improve model performance and computational efficiency. We report computational results using 2022 mobile phone data from placer.ai with 37 residential zip codes and 149 community pharmacies in Louisville, KY. Future work will extend the study with block group data in ArcGIS.

WA60

Building B Level 3 B319

Game Theory Applications

Contributed Session

Contributed

Chair: Ceyhun Eksin, Texas A&M University, College Station, TX, United States

1 - A Game-Theoretic Framework for Multi-Player Global Lithium Supply Chain

Sarasadat Alavi, University of Texas at Arlington, Arlington, TX, United States

Lithium, a critical mineral essential for clean energy technologies, has prompted numerous nations and stakeholders to invest in domestic supply chain development to achieve resource independence. However, the complex interplay of tariffs, policies, production costs, and pricing strategies across global markets creates interdependencies that significantly impact individual national optimization efforts. While previous studies have explored optimization models for the global lithium supply chain, limited attention has been given to the strategic interactions among countries and their mutual influence on supply chain design and performance.

This study introduces a novel multi-level optimization framework that captures both strategic and operational interactions within and between multiple decision-making entities (i.e., countries). The proposed framework employs a game-theoretic approach to model the hierarchical nature of decision-making, where each country seeks to optimize its own lithium supply chain while anticipating the responses of others. The focus of this presentation is on formulating a multi-level optimization model and implementing exact solution approaches based on decomposition methods to illustrate causal relationships between nations and quantify their mutual effects on supply chain performance and economic outcomes. Our findings provide valuable insights for policymakers seeking to navigate the complex trade-offs between national resource security and economic efficiency in the global lithium market.

2 - Vector-Valued Games: Characterization of Equilibria in Matrix Games

Giovanni Crespi, Università Cattaneo LIUC, Castellanza, Italy, Daishi Kuroiwa, Matteo Rocca

In this work, we investigate equilibrium concepts for games with vector-valued payoffs, focusing on Nash, weak Nash, and proper Nash equilibria. We provide a general framework for defining these equilibria using cone-based vector orderings, and we explore the relationships among them. Existence results are established through fixed-point theorems and scalarization techniques, allowing us to represent vector preferences via weight vectors and scalar-valued functions.

In the special case of vector-valued matrix games, we show that proper and classical Nash equilibria coincide, and we provide existence results for both Nash and weak Nash equilibria. We further characterize best response correspondences by constructing functions from the payoff matrices and the players' preference cones, showing how these can be used to identify all equilibrium points.

For two-player, two-strategy, two-objective matrix games, we develop a complete classification of best correspondence sets using convex analysis and scalarization. This leads to an explicit description of all possible equilibrium regions in the mixed strategy space.

Our approach provides a systematic method for analyzing multi-objective strategic interactions, with applications ranging from environmental management to negotiation and public policy. The results offer theoretical insight and computational tools for identifying equilibrium behavior in settings where players face multiple, possibly conflicting, objectives.

3 - Solving Markov Games with Coupled Local Constraints via Dual Gradient Descent

Ceyhun Eksin, Texas A&M University, College Station, TX, United States, Soham Das, Santiago Paternain, Luiz Chamon

We introduce a dual gradient descent method to solve Markov games subject to local but coupled constraints. These constrained Markov games (CMGs) model multiagent systems operating in dynamic environments where each agent seeks to maximize its individual utility while respecting specific local constraints. These constraints, which capture real-world limitations such as resource budgets or safety requirements, are local in the sense that they pertain to individual agents but are coupled through dependencies on other agents' actions and the shared environment state. Our approach builds on the concept of the unconstrained Lagrangian game, which agents repeatedly solve based on the current values of their own local Lagrange multipliers. Agents simulate trajectories to estimate cumulative rewards and constraint violations, then adjust the multipliers using a dual gradient step grounded in their local observations. This primal-dual process produces a sequence of policy profiles that converge to a nonstationary and feasible Nash equilibrium of the original constrained game. This work extends previous frameworks for solving Markov games with global constraints, providing a principled approach to safe multiagent learning under localized, interdependent feasibility conditions.

WA63

Building B Level 4 B403

AI and Society

Contributed Session

Contributed

Chair: Md Abdul Quddus, North Carolina State University, Cary, NC, United States

1 - A Similarity Based Method Against ChatGPT Based Plagiarism

Petros Xanthopoulos, Stetson University, DeLand, FL, United States, Konstantinos Xylogiannopoulos, Panagiotis Karampelas, Georgios Bakamitsos

Large language models (LLMs) have found many profound applications and for many are necessary tool that boosts productivity. At the same time they enable several malicious applications with plagiarism and copyright infringement being one of them. In this presentation we show examples of ChatGPT paraphrased news articles, we capture the inability of AI detectors to detect these paraphrases and we propose a similarity based approach for tackling this emerging problem. We demonstrate our method with a database of real news articles from a major international news vendor. Our proposed method achieves performance of over a set of characteristic classification performance metrics such as accuracy, sensitivity, specificity and F1 statistic.

2 - Collaboration and Knowledge Sharing in Teams: a Graph-Based Deep Learning Approach

Qianyin Xia, Georgia State University, Atlanta, GA, United States, Yi Zhao, Sarang Sunder

Collaboration and knowledge sharing are foundational to modern organizations, especially in environments where teamwork drives performance. However, understanding how individuals contribute to and learn from multiple teams remains empirically and methodologically challenging. This paper introduces a Two-Stage Hypergraph Neural Network (TSHyGNN) framework to model agent behavior in multi-agent, multi-team settings. Leveraging hypergraph structures, the model captures both within-team collaboration and cross-team knowledge transfer. To address endogeneity in team formation, we adopt a two-stage estimation approach: the first-stage hypergraph model predicts team formation and produces a correction term, which is incorporated into the second-stage performance model to account for selection bias. Using salesforce data from a large financial services firm, we show that TSHyGNN outperforms traditional econometric and network-based models in predicting team outcomes. Simulation results suggest that replacing a novice with an experienced salesperson can generate approximately \$10K in direct team collaboration gains and \$16K from knowledge spillovers across teams, with sales increasing by 1-3%. The deep learning architecture is designed to be interpretable by aligning its components with collaboration and knowledge-sharing processes. This approach offers a scalable and flexible tool for team design, with broad applicability to settings characterized by overlapping team structures and dynamic collaboration networks.

3 - Syngene: a Knowledge Graph Framework for Root Cause Analysis in Regulated Life Sciences Manufacturing

Darshan Desai, Quantic School of Business and Technology, Washington DC, DC, United States, Manashri P. Vyas, Ashish Desai

Timely and accurate root cause analysis (RCA) is essential in life sciences-related manufacturing industries such as pharmaceuticals and medical devices, where quality deviations can lead to costly recalls, regulatory scrutiny, and public health risks. Yet RCA processes remain largely manual, relying on siloed systems and time-intensive expert investigations, making it difficult to maintain traceability, audit readiness, and fast decision-making when it matters most.

To help close this gap, we introduce a data-centric framework that combines knowledge graph traversal and rule-guided reasoning to automate RCA. The system, called SynGene, constructs a process graph based on ISA-88/95 principles, linking batches, materials, equipment, operators, lab tests, and deviations into a cohesive structure. SynGene's traversal agent identifies causal pathways through multi-hop queries, surfacing contamination risks and compliance issues with high interpretability. Unlike black-box approaches, it builds on logic-driven methods and domain-specific ontologies to provide transparent, auditable insights.

Evaluation on a synthetic benchmark designed to reflect common batch genealogy structures and deviation types in life sciences manufacturing showed that SynGene significantly reduced RCA effort and response time in simulated quality incident scenarios. The traversal agent identified failure pathways in under 10 minutes, compared to manual trace-style queries that required substantially more time and domain knowledge. The system accurately identified affected batches across multi-hop failure chains, and its visual outputs including failure propagation paths and node clusters enhanced interpretability and supported audit readiness. This work illustrates how graph-based, domain-aligned representations can advance decision automation in regulated industries, with broad applicability across life sciences manufacturing.

4 - Inducing Interpretability In Deep Learning via Scalable, Kernel-Based Density Estimation

Ayyuce Begum Bektas, Sloan Kettering Institute, New York, NY, United States, Mithat Gonen

Interpretability remains a central concern in deep learning applications across healthcare, finance, and autonomous systems where transparency is essential. We introduce a scalable method that incorporates random Fourier features with softmax-weighted kernel density estimation into a unified learning objective. This construction enables efficient approximation of kernel evaluations while emphasizing structured density modeling. We conduct a sensitivity analysis on the dimensionality of the mapped space to characterize its effect on computational cost. The framework also allows the integration of multiple kernels within a single model, enabling distinct treatment of heterogeneous feature sets such as gene groups or modality-specific inputs. The method is evaluated in a conditional density estimation setting using real world data, demonstrating competitive predictive performance alongside interpretable behavior.

5 - An AI Approach for Learning the Spectrum of the Laplace-Beltrami Operator

Yulin An, Penn State University, State College, PA, United States, Enrique Del Castillo

The spectrum of the Laplace-Beltrami (LB) operator is central in geometric deep learning tasks, capturing intrinsic properties of the shape of the object under consideration. The best established method for its estimation, from a triangulated mesh of the object, is based on the Finite Element Method (FEM), and computes the top k LB eigenvalues with a complexity of $O(Nk)$, where N is the number of points. This can render the FEM method inefficient when repeatedly applied to databases of CAD mechanical parts, or in quality control applications where part metrology is acquired as large meshes and decisions about the quality of each part are needed quickly and frequently. As a solution to this problem, we present a geometric deep learning framework to predict the LB spectrum efficiently given the CAD mesh of a part, achieving significant computational savings without sacrificing accuracy, demonstrating that the LB spectrum is learnable. The proposed Graph Neural Network architecture uses a rich set of part mesh features —including Gaussian curvature, mean curvature, and principal curvatures. In addition to our trained network, we make available, for repeatability, a large curated dataset of real-world mechanical CAD models derived from the ABC dataset (Koch et al., 2019) used for training and testing. Experimental results show that our method reduces computation time of the LB spectrum by approximately 5 times over linear FEM while delivering competitive accuracy.

6 - Toward Precise Sleep Apnea Screening: a Sound-Based Approach via Temporal Anchor Contrastive Learning and Synergistic Distillation

Fulin Cai, Arizona State University, Tempe, AZ, United States

Obstructive sleep apnea (OSA) is a common sleep disorder associated with serious cardiovascular and metabolic risks, yet it remains widely underdiagnosed due to the cost and inconvenience of in-lab polysomnography. Recent studies have explored deep learning (DL)-based screening using ambient audio recordings as a contactless and scalable alternative. However, existing segment-based detection models struggle to precisely delineate apnea event boundaries, which is essential for estimating apnea duration—an increasingly recognized clinical biomarker.

Accurate delineation of OSA events requires models that can address two core challenges: temporal variation in respiration patterns and ambiguous transitions between respiratory states, where changes occur gradually rather than abruptly. To address these issues, we propose TACO-SynD (Temporal Anchor Contrastive learning with Synergistic Distillation), an auxiliary DL module for second-level OSA detection using sleep sound recordings. Built on a second-level classification model, TACO-SynD introduces: (1) temporal anchor contrastive learning, which aligns latent representations to context-specific breathing patterns, and (2) synergistic distillation, which transfers localized relational knowledge to enhance the global classifier.

A comprehensive evaluation is conducted on the real-world PSG-Audio dataset across both OSA screening and event detection tasks. TACO-SynD achieves state-of-the-art screening performance with mean absolute errors of 6.41 for AHI and 1.33 for MAD, along with superior event-level alignment and boundary localization. These results demonstrate the framework's robustness and potential as a scalable solution for in-home OSA monitoring.

7 - An Image-Based Deep Learning Approach for Precise Sorting of Knitted and Woven Textiles

Md Abdul Quddus, North Carolina State University, Raleigh, NC, United States, Yihan Huang, Karen Leonas

The rapid growth of the textile industry has led to a significant increase in textile waste, posing serious environmental and resource challenges. Traditional sorting methods are often labor-intensive, slow, inaccurate, and difficult to scale sustainably. In contrast, computer vision and deep learning techniques offer a promising alternative by enabling automatic feature extraction from complex data, achieving high classification accuracy and supporting scalable, real-time sorting using spectral and visual inputs. Our research aims to develop a practical, production-ready solution for textile recycling facilities by leveraging real-world image data collected from both high-resolution scanners and mobile phone cameras. To enhance generalization across these heterogeneous data sources, we will develop an advanced deep learning architecture customized for fabric structure classification. This approach allows models to be trained collaboratively across decentralized devices without the need to pool sensitive or device-specific data, making the system more robust to variations in lighting, resolution, and device quality. By combining advanced AI algorithms, textile domain expertise, and a diverse image dataset, we will create a reliable and efficient system to classify fabric structures such as Plain Weave, Jersey Knit, Rib Knit, Tricot Knit, and related variants with high accuracy. The outcome will be a scalable, intelligent fabric-sorting system that addresses a critical bottleneck in textile recycling, improves material recovery quality, and promotes a more sustainable circular economy.

WA64

Building B Level 4 B404

Innovation, Incentives, and Behavior in Platform Ecosystems

Contributed Session

Contributed

Chair: Lahiru Gunasekara, Lincoln University, Lincoln, 7647, New Zealand

1 - Strategic Data Acquisition and Pricing Decisions in Platform Competition with Data Broker Collaboration

Jingjing Li, Tianjin University, Tianjin, China, People's Republic of, Haiyang Feng

As a tradable and strategic asset in the digital economy, data has driven the rapid development of a thriving data brokerage industry. The rapid emergence of the data brokerage market has fundamentally reshaped how firms compete and create value. Data brokers aggregate and analyze vast datasets, which are then distributed to client firms for targeted advertising and personalized pricing. Collaborating with data brokers has become a strategic lever for firms seeking competitive advantages in the advertising market. However, this practice also raises significant privacy concerns while simultaneously intensifying market competition among firms. To investigate firms' decisions regarding data acquisition and service pricing, this study develops a game-theoretic model that considers two horizontally differentiated firms. The results demonstrate that the firm with superior data analytics capability reduces data acquisitions as the unit price of data increases, while the firm with weaker capability exhibit a non-monotonic purchasing behavior—initially decreasing and then increasing their data demand as data price escalates. A counterintuitive finding reveals that when advertising revenue per user is high and unit price of data is relatively low, the firm with weaker analytics capability tends to set a higher price. Furthermore, under higher data purchasing price, the firm with stronger analytics capability purchase larger volumes of data, attract more users, and achieve higher profits compared to its less analytically capable counterpart. Conversely, in instances of low data price, a scenario of market dominance by stronger analytics capability company emerges, leading to a winner-takes-all outcome.

2 - A Framework for Identifying Causal Structures of AI Service Failures Using Text Mining and Network Analysis

Su-Yeon Kim, Incheon National University, Incheon, Korea, Republic of, Do-Hyeon Ryu

Artificial intelligence-based services (AIS) use autonomous learning and adaptive capabilities to predict user behavior and deliver personalized experiences. The increasing use of AIS coincides with heightened awareness of unintended social, ethical, and operational risks. These risks often materialize as service failures, such as biased outputs, or security breaches. Prior research has focused on categorizing failure types, providing limited understanding of the causal structures underlying the emergence and propagation of such failures. Without this causal insight, organizations are unable to trace causes, resulting in repeated failures or inefficient resource allocation.

This study presents a framework to identify causal elements and their interrelations underlying AIS failures. Based on the Failure Mode and Effects Analysis (FMEA), a three-step structure cause – failure – effect was adopted to guide the extraction process. Failure types were defined from prior literature into four categories: function, security, information, and interaction. Over 15,000 news articles were embedded using Sentence-BERT and classified accordingly. For each failure type, a multi-turn Question Answering model was applied to sequentially extract failures, causes, and effects using tailored queries designed in accordance. The extracted elements were assembled into a causal network, where analysis identified high-centrality nodes, major propagation paths, and clustered failure patterns.

The analysis revealed that AIS failures commonly result from complex interactions among multiple factors rather than singular causes. The proposed framework reveals latent causal structures and identifies high impact paths and key triggers. This insight enables organizations to anticipate how failures propagate, prioritize points of

intervention, and develop targeted mitigation strategies.

3 - Digital Consumption and Self Control

Zheng Wu, Purdue University, West Lafayette, IN, United States, Jiong Sun

The rise of digital vice goods—such as social media, streaming platforms, and video games—has heightened concerns over consumer self-control, as these services often promote impulsive and habitual use. Excessive engagement is linked to negative mental health outcomes, prompting policymakers to advocate for pre-consumption planning as a behavioral intervention. Our study develops a theoretical model to examine how firms price digital vice goods using subscription versus pay-per-use schemes, accounting for consumers with present-biased preferences. We analyze how these pricing strategies interact with product harm and consumer self-control to shape firm profit, consumer welfare, and public health outcomes. Our findings show that pricing digital vice goods differs markedly from standard goods due to behavioral frictions. Without pre-planning, firm profits decline with rising harm. However, when consumers engage in pre-planning, firms can benefit from market segmentation. Interestingly, while pre-planning is often seen as consumer-friendly, we find that some adopting pre-planning may reduce consumer surplus and worsen public health. These results offer practical insights for firms, consumers, and policymakers seeking to design effective pricing strategies and interventions in digital markets.

4 - Gig OR Employee? Platform's Joint Staffing and Pricing Decisions Under Hybrid-Staffing Mode

Xuyan Xin, University of California, Los Angeles, Los Angeles, CA, United States

In the rapidly evolving landscape of on-demand services, platforms are increasingly adopting a hybrid-staffing mode that combines full-time employees with part-time gig workers. This approach leverages the stability of full-time staff and the flexibility of gig workers, enabling platforms to efficiently meet fluctuating consumer demands. To identify optimal staffing and pricing strategies, we develop a two-period model analyzing these joint decisions. Our findings highlight several key insights. First, the platform's staffing choice is influenced by employee salary costs and the labor pool size of gig workers. When both factors are moderate, a hybrid-staffing mode is optimal. Additionally, as the frequency and intensity of demand surges increase, platforms are more inclined to adopt a hybrid staffing approach. Second, within the hybrid staffing framework, gig workers can fulfill multiple roles. They not only act as supplementary resources during peak hours but also maintain operational flexibility in both high and low demand states in some cases. This dual role necessitates adaptive hiring and pricing strategies, which may exhibit non-monotonic patterns in response to market changes. Third, platforms consistently benefit from hybrid-staffing, achieving higher profits compared to single-staffing mode. Moreover, this mode enhances labor and consumer welfare, consistently outperforming gig-only approach. Our study offers valuable insights for platforms to optimize hybrid staffing and pricing strategies, emphasizing the impact on stakeholder welfare within the ecosystem.

5 - Do Subsidies Shape Channels? the Role of Differentiated Incentives in New Energy Vehicle Channel Design

HELI ZHANG, Institute of Operations Research and Analytics, National University of Singapore, Singapore, Singapore, Yexia Zhang, Cejun Cao, Zhongzhu Chu, Mabel C. Chou

Given the global popularity of new energy vehicles (NEVs), a paramount problem for the NEV manufacturer is to consider the selling format and decide whether to encroach on the reseller by setting up a direct channel under differentiated government subsidies. In this paper, we first empirically examine the relationship between government subsidies and NEV manufacturers' channel strategies based on data from the Chinese NEV market. Then, a game-theoretical model is established to investigate these interactions between different channel structures and three types of government subsidies: no subsidy program (NSP), sale-price subsidy program (SSP), and technology subsidy program (TSP), considering heterogeneous consumer preferences. Empirical results reveal that the TSP significantly reduces the number of direct-selling stores but has no significant impact on the number of resellers, whereas the SSP significantly reduces the number of resellers without impacting direct-selling stores. Theoretical findings indicate that prices under the SSP are higher than those under the NSP and TSP, and SSP enhances profits for both manufacturers and retailers when only reselling channels exist. In the TSP, encroachment is favorable for manufacturer only when cross-price competition degree is low, and the TSP proves more effective in increasing social welfare when only a direct channel exists. Both SSP and TSP can stimulate consumer demand to an equal extent but operate through different mechanisms. Our study provides useful insights for the NEV manufacturer to make channel distribution choices and for the government to design subsidy schemes.

6 - Kick Out Innovation: Explore the Conditions Where Innovation Cannot Invade in Competitive Market

Songhan Wu, University of Illinois Urbana-Champaign, Urbana, IL, United States

The introduction of radical innovations frequently disrupts existing markets, yet the mechanisms governing their adoption remain complex and multifaceted. While prior research has primarily focused on binary adoption decisions within monopolistic settings, this study extends the framework by examining how social comparison influences the diffusion of radical innovations in competitive markets with multiple incumbents. Drawing on evolutionary game theory, we model consumer decision-making under ahead-seeking and behind-averse behaviors, revealing the conditions under which innovations succeed or fail. Unlike findings in monopolistic contexts, our results reveal that excessive ahead-seeking behavior can suppress innovation, whereas moderate behind-averse behavior can drive convergence toward a particular product (probably the compromise one). These insights challenge conventional wisdom on social pressure and innovation diffusion, offering a fresh perspective on how competitive structure and psychological effects shape market dynamics.

7 - Optimising Retailer-Level Take-Back Systems for Low- and Medium-Value Used Products for a Circular Economy

Lahiru Gunasekara, Lincoln University, Lincoln, New Zealand, David Robb

Used product take-back plays a significant role in the transition to a Circular Economy. Many businesses are introducing retailer-level take-back schemes for low and medium-value products by deploying collection bins for customers to drop off used products. However, the literature has not given much attention to these schemes compared to those with high recovery value products. The viability of such schemes depends on several factors, including the volume of returns and the contamination in the collection bins. We consider a two-echelon reverse supply chain consisting of a retailer taking back low and medium-value used products and delivering them to a recycler for value recovery. We mathematically model the optimal retailer decisions, including the return incentive paid to customers, investment in contamination reduction, the number and capacity of bins, their product assignment, and time between shipments from the retailer to the recycler, and apply the optimisation model to a case study. In addition, our numerical experiments with globally representative parameter value combinations provide generalisable insights for theory and practice. For example, recycling low-value products like plastic containers is economically infeasible for retailers under all conditions. However, including at least one medium-value product, such as apparel, in a take-back programme aids in offsetting its costs. Investing in reducing contamination for any product type is financially unwarranted unless the costs of separation and disposal exceed that of contamination reduction investment. Furthermore, combining multiple product streams into fewer bins could improve financial results, but keeping bins separate may offer psychological benefits to users.

WA65

Building B Level 4 B405

AI-Driven Innovation in Supply Chain and Business Process

Invited Session

Artificial Intelligence

Chair: Huihui Chi, NEOMA business school, Mont-Saint-Aignan, France

Co-Chair: Jaecheol Park, Nanyang Technological University, Singapore, N/A, Singapore

1 - A Framework for Human-AI Collaboration for Theory Building

Ke Li, INSEAD, Fontainebleau, France, Philip Parker, Phanish Puranam, Eric Uhlmann, Spyros Zoumpoulis

We propose a novel human-AI collaborative framework for theory building. Human researchers are adept at generating creative hypotheses but are prone to biases. Machine learning (ML) models can identify robust empirical patterns but lack explanatory reasoning. Recent advances in generative AI (GenAI) offer a third capability:

the ability to simulate abductive reasoning and generate hypotheses—though they risk producing hallucinations.

Our framework integrates the strengths of human judgment, ML-driven pattern discovery, and GenAI-generated explanation through a structured process. Focusing on inequality discourses (e.g., gender, race, and socioeconomic disparities), we leverage both human intelligence and AI to examine how these inequality discourses spread across different historical periods, societies, and social sectors.

We address two core research questions: (1) How accurately can humans, compared with GenAI models, generate hypotheses about empirical patterns? and (2) How can collaboration between humans, ML, and GenAI improve theory building?

We present results from a pilot forecasting survey comparing the accuracy of individual and aggregated human forecasts to the accuracy of state-of-the-art GenAI models, drawing insights into the value of GenAI and collective human intelligence for theory building. We argue that human–AI collaboration offers a potential solution to challenges in science, including theory fragmentation and the replicability crisis.

2 - Designing for Designers: a Multimodal Hypergraph Rag System to Enhance Automotive Design

Xiaoke Zhang, University of British Columbia, Vancouver, BC, Canada, Mi Zhou, Gene Moo Lee, Angela Kwon

The growing adoption of large language models (LLMs) across industries highlights the need for domain-specific systems that leverage an organization's proprietary knowledge. In the automotive sector, general-purpose LLMs often lack specialized expertise and may produce irrelevant or misleading outputs, hindering vehicle designers' creative processes. To address these challenges, we partner with Kia, a leading automotive manufacturer in South Korea, to develop a designer-oriented, multimodal hypergraph retrieval-augmented generation framework for vehicle concept ideation. Our framework consists of two core components. First, we construct a custom hypergraph knowledge base that captures complex relationships between customer feedback (text modality) and design assets (visual modality). Second, we design an interactive chatbot interface that accepts both free-form text and image inputs, retrieves relevant subgraphs from the knowledge base, and generates contextually grounded responses. We plan to evaluate the prototype through randomized online experiments and user studies involving Kia's design teams. This work will contribute to design science by proposing a scalable method for multimodal knowledge representation and demonstrating how interactive AI tools can support domain-specific creative exploration.

3 - How Does AI Change Drug Development? Evidence from Clinical Trial Phases and Drug Types

Angela Kwon, University of British Columbia, Vancouver, BC, Canada, Jaecheol Park, Gene Moo Lee

We examine how the AI capabilities of pharmaceutical firms affect their drug development processes, with a focus on clinical trials. Clinical trials test the safety and efficacy of drug candidates on human subjects across three phases. While the first two phases (Phases I & II) assess preliminary safety and efficacy in smaller populations, the last phase (Phase III) investigates thorough effectiveness in larger populations prior to post-market surveillance. To examine the effect of the AI capabilities of firms on clinical trial outcomes, we leverage job postings data and clinical trial records provided by the U.S. National Library of Medicine. Our findings suggest that AI capabilities have significant effects on the quantity of clinical trial outcomes (total trials, new trials, and retrials). To further explore underlying mechanisms, we delve into detailed trial process outcomes: (1) trial duration as a proxy for the pace of clinical progress; (2) inter-phase transition time for the preparation between phases; and (3) success rate for the probability that a drug candidate progresses to the next phase. Moreover, we find the heterogeneous effects across drug types and trial phases. This study contributes to the healthcare IS literature by empirically demonstrating the business value of AI in enhancing drug development processes. It also provides practical guidance on AI investment and implementation for pharmaceutical firms, regulatory agencies, and policymakers.

Keywords: Artificial intelligence, AI capability, AI innovation, clinical trials, biotech and pharmaceutical industries

WA66

Building B Level 4 B406

Data Mining for Healthcare and Society

Contributed Session

Contributed

Chair: Neda Amiri Rad, Binghamton University, State University of New York, 2 Crandall street, Binghamton, 13905, United States

1 - Magic: a Multi-Task Gaussian Process for Joint Imputation and Classification in Healthcare TIME Series

Dohyun Ku, Georgia Institute of Technology, Atlanta, GA, United States, Catherine Chong, Visar Berisha, Todd Schwedt, Jing Li

Time series analysis in healthcare has become instrumental in improving both care quality and predictive analytics. The ability to forecast disease progression and mortality risks can offer substantial life-saving benefits to patients. However, healthcare time series data often presents two critical challenges: time misalignment and data sparsity. Traditional approaches address these challenges through a two-step process: first imputing missing values, then predicting the response variable. The primary limitation lies in its inability to leverage class label information during imputation and its independent parameter optimization. We present the **Multi-task Gaussian Process for joint Imputation and Classification (MAGIC)**, a novel framework that simultaneously performs missing value imputation and class label prediction. Our innovative approach leverages class label information to enhance imputation accuracy through Multi-task Gaussian Process model, which in turn improves the reliability of subsequent predictions. By integrating imputation and classification within a single framework, MAGIC enables real-time predictions of both future variable trajectories and disease progression patterns. This capability facilitates early disease detection and proactive strategies that can significantly impact patient care decisions and treatment outcomes. We validate the model through two healthcare applications: 1) prediction of post-traumatic headache after mild traumatic brain injury, and 2) prediction of in-hospital mortality within 48 hours of ICU admission. In both applications, MAGIC demonstrated superior predictive accuracy compared to existing methods.

2 - Constructing Optimal Patent Datasets for An Effective Patent Analysis: a Multi-Objective Optimization Framework

Doyoung Park, Seoul National University, Seoul, Korea, Republic of, Sungjoo Lee

As data-driven technology management is getting increasing attention, patent data has been recognized as a critical source of technological knowledge. While the effectiveness of patent analysis largely depends on how well the dataset is constructed prior to use, little attention has been paid to validating the quality of the collected datasets in terms of the objective of analysis. In this study, we focus on technology trend analysis, technology opportunity analysis, and patent quality analysis as the objectives of patent analysis. Then, we aim to propose a dataset construction framework with the aim of evaluating and filtering patent data based on different objectives of analysis. To achieve the aim, a domain-specific filtering process is first applied to eliminate noise to ensure alignment between the collected patents and the intended analysis domain. Second, key components that shape the objectives of each patent analysis are identified. Third, a multi-objective optimization process is employed to maximize or minimize the identified components based on different analysis objectives. The optimization problem leverages patent metadata including citation counts, assignee information, and family size, and identifies efficient frontiers among the different objectives. By evaluating and refining patent datasets, the proposed approach can enhance the relevance, reliability, and strategic value of patent analysis.

3 - Considering Overdiagnosis in Cancer During Computer-Aided Diagnosis Decision Making

William Brown, Louisiana State University, Baton Rouge, LA, United States, Evangelos Triantaphyllou

Overdiagnosis is an emerging area of concern within healthcare. Overdiagnosis is when an individual is diagnosed with a disease, but they will never be harmed or have symptoms of it, which causes unnecessary stress. Such a disease is detected through asymptomatic screening, which is becoming increasingly popular with the use of Computer-Aided Diagnosis (CAD), especially for cancers such as breast, prostate, and lung cancer. Overdiagnosed cases are not false-positives because the affected

patients have the disease, but it may be an indolent or slow-growing disease. It is not possible to distinguish between overdiagnosis and successful treatment currently, thus the prevalence of overdiagnosis inflates the success rate of treatments. Overdiagnosis can be observed in a population screened for cancer because cancer is usually fatal if untreated. Therefore, if a cancer screening program is introduced to a population and there is a large increase in the incidence rate, but a relatively stable or slightly reduced mortality rate over an extended time, then it indicates that overdiagnosis has occurred. This phenomenon is often mentioned in medical literature along with CAD. However, although CAD is frequently mentioned in computer science and engineering literature, overdiagnosis is scarcely mentioned. This is concerning because CAD is increasingly used for asymptomatic screening. CAD developers need to begin to implement data mining approaches to offer prognostic information for detected cancers, along with the binary classification of cancer. Such prognostic information can help healthcare professionals and patients make better decisions about treatment options, which can reduce the harm of overdiagnosis.

4 - Advancing Heart Failure Readmission Prediction Through Feature Augmentation and Stratified Machine Learning

Sonia Jahangiri, Rochester Institute of Technology, Rochester, NY, United States, Nasibeh Azadeh-Fard

Reducing hospital readmissions for heart failure (HF) remains a national healthcare priority. While machine learning (ML) models have demonstrated promise in predicting HF readmissions, challenges persist in enhancing model accuracy and ensuring equity across patient subgroups. This study aims to improve HF readmission prediction by incorporating additional clinically relevant features and conducting stratified analyses. We extend an existing ML framework by engineering new features, such as procedure codes and other underutilized variables, to capture more nuanced aspects of patient care. Statistical methods will be employed to identify significant predictors, which will then inform a heuristic feature selection process. A range of ML algorithms will be applied to evaluate the impact of the expanded feature set on prediction performance. To further explore disparities, stratified models will be developed for key subgroups, including HF subtypes and gender. These subgroup analyses are designed to uncover distinct risk profiles and improve the personalization of readmission prediction. The anticipated outcome is a more robust, generalizable, and equitable prediction framework that supports data-driven interventions for at risk patients.

5 - Incremental Learning with Multi-Group Data Stream for Diabetic Retinopathy Prediction

Enrico Laoh, Oklahoma State University, Stillwater, OK, United States, Tieming Liu

Diabetic retinopathy (DR), a leading cause of blindness, disproportionately affects rural U.S. populations due to limited access and low compliance with annual eye exams. Many patients remain undiagnosed as DR shows no early symptoms, despite underlying damage. While AI-based image screening aids early detection, it remains costly and inaccessible for many. This highlights the need for a non-image-based DR screening model using routine clinical data—already available at no extra cost—to enable wider, earlier, and more equitable detection. The Incremental Learning with Multi-Group Data (IL-MGD) framework is proposed in this study to lower the computational burden and completely retrain the model when new data becomes available. The IL-MGD dynamically integrates continuous data updates, enhancing its robustness and relevance. IL-MGD robustness was tested in this study, and a thousand data subsets were generated nationwide from hospitals in the Cerner Real World Data (CRWD) Warehouse. The result demonstrated high accuracy and robust performance of the Gaussian naïve Bayes model, tested under a thousand model-building trials including different dataset variations, with mean accuracy rates of 97.08% and F1-scores around 97.98%, reduced only around 2% from the best offline neural-based model could achieve. With the performance metrics standard deviation below 1%, the framework shows consistency in the prediction. In addition, the framework allows the model to build in under 90 seconds on average, which is around a 94% time reduction compared to the fastest online neural-based model could achieve. These notable potentials make the IL-MGD potentially a scalable and sustainable tool for early DR detection.

6 - Profiling Depression Risk Groups Among Students Using Pca and Cluster Analysis: Identifying Hidden Risk Patterns in Lifestyle and Academic Stress Factors

Neda Amirirad, Binghamton University, Binghamton, NY, United States, Suxia Lu

Depression among university students is a complex issue that demands systematic, data-driven methods to uncover underlying risk groups. In this study, we analyzed survey data from 12,690 students using Principal Component Analysis (PCA) for dimensionality reduction, followed by K-Means clustering to identify distinct subpopulations. Nine psychosocial and behavioral variables were included, such as academic pressure, satisfaction, financial stress, sleep duration, and dietary habits. PCA reduced the data to six principal components, capturing approximately 88% of the total variance. Cluster validation through silhouette scores and the Elbow method indicated that a three-cluster solution was optimal. The identified groups were: (1) a *High-Risk* cluster, characterized by high academic and financial stress and low satisfaction; (2) a *Low-Risk* cluster, with balanced lifestyle behaviors and positive mental health indicators; and (3) a *Hidden Risk* cluster—students with strong academic performance but poor sleep and dietary habits. These results highlight that academic success does not necessarily equate to psychological well-being. Our findings demonstrate the utility of unsupervised learning techniques in revealing latent risk profiles and can inform targeted mental health interventions for student populations.

WA67

Building B Level 4 B407

Real-World Data Analytics and Optimization for Improving Power Grid Operation, Planning, and Resilience

Invited Session

Data Mining

Chair: Zhaoyu Wang, Iowa State University, 2520 Osborn Drive, Ames, IA, 50011, United States

1 - Data-Driven Methods for Improving Distribution System Planning and Operations

Joe Azzolini, Sandia National Labs, Albuquerque, NM, United States

Electric power distribution grids are undergoing significant changes due to increased electrification and the deployment of distributed generation. Conventional analysis methods require distribution system planners and operators to conduct time-consuming simulations on detailed grid models, which are often created and maintained manually. Fortunately, there has been a steady increase in the number of measurement devices on the grid—from residential smart meters to intelligent switches on the medium-voltage system—that generate data at much finer temporal and spatial resolutions. This presentation will discuss how we leverage this data to provide distribution engineers with data-driven alternatives to conventional model-based analyses.

2 - TBD

Rui Yang, National Renewable Energy Lab, Denver, CO, United States

3 - Proactive Planning for Distribution System Resilience Under Uncertainty

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

Distribution systems face growing threats from extreme events such as wildfires, requiring decision frameworks that balance risk mitigation with operational costs. In this talk, I present two recent projects. The first analyzes wildfire resilience interventions—Public Safety Power Shutoffs (PSPS) and Enhanced Powerline Safety Settings (EPSS)—quantifying their effectiveness in reducing ignition risks and their social and operational trade-offs. The second develops an adaptive robust optimization framework that jointly considers proactive measures (e.g., hardening, de-energization) and reactive actions (e.g., restoration), adapting uncertainty sets to evolving conditions. Together, these approaches highlight how robust, adaptive, and socially responsible strategies can enhance grid resilience under deep uncertainty.

4 - Revenue Adequacy-Constrained Optimal Transmission Switching: a Non-Iterative Milp Formulation

Bai Cui, Iowa State University, Ames, IA, United States

Optimal Transmission Switching (OTS) provides a promising approach to enhancing grid efficiency and mitigating congestion while challenging revenue adequacy (RA) in Financial Transmission Rights (FTR) markets, as traditional Simultaneous Feasibility Tests (SFT) rely on static network topology assumptions. To address the optimization of the original model incorporating line-switching variables and its dual variables under RA constraints, a RA-constrained optimal power flow model, referred to the OTS-based optimal price flow formulation (OTS-OPrF), is developed to enable the co-optimization of network topology and generation. By integrating binary transmission variables into the Karush-Kuhn-Tucker (KKT) conditions, the model reformulates Locational Marginal Prices (LMPs) and converts the problem into a

unified Mixed-Integer Linear Programming (MILP) formulation. The behavior of LMPs parallels DC circuit laws, and differences in LMPs can be captured through information on network topology, line susceptance, and the identification of congested lines. This approach eliminates the need for the iterative-based method, significantly reducing computational burden while maintaining solution precision. Case studies on IEEE 39-bus, 118-bus, and 300-bus systems demonstrate the effectiveness of the model, with computational efficiency improvements ranging from 45.33% to 93.80% compared to iterative methods.

WA68

Building B Level 4 B408

Energy Systems Optimization and Planning

Contributed Session

Contributed

Chair: Brent Austgen, Sandia National Laboratories, Lowell, IN, United States

1 - Web-Based User Interfaces for Mathematical Models to Support Decisions: the Fuel Distribution Network Optimizer

Walid Matar, King Abdullah Petroleum Studies and Research Center, Riyadh, Saudi Arabia, Pavithra Shetty, Abeer Aghamdi, Nourah Alhosain, Jaidaa Alidrisi, Amr Alruthai, Abdelrahman Muhsen, Lama Yaseen

Mathematical models provide quantitative analyses that can inform policy decisions. We argue that two attributes can support models' adoption by policymakers: an intuitive graphical interface and active engagement between modelers and policymakers. Graphical interfaces make models accessible to non-specialists. Web-based implementations of such interfaces are preferred because they simplify the deployment of updates to the underlying software. Likewise, they minimize disruptions to the end-users when model updates are deployed. Also, active engagement between modelers and policymakers is essential to build trust. End-users may be involved in different stages of model and interface development. We exemplify these attributes in an optimization model and web-based interface framework we developed to support decisions pertaining to the refined fuels' distribution in Saudi Arabia. The model's end-users were involved during model parameterization, advised on model constraints, and helped with developing the inputs and outputs screens of the interface. They were also crucial during the model validation process.

2 - Optimal Equipment Layout Strategy for Small Modular Reactors Using Mixed-Integer Programming

HyungJu Kim, Korea University, Seoul, Korea, Republic of, Taesu Cheong

This study aims to analyze and optimize the efficient layout of small modular reactors (SMRs) and micro modular reactors (MMRs) from an industrial engineering perspective. As the demand for carbon-neutral and decentralized energy supply systems grows, SMRs and MMRs, which are flexible and small-scale nuclear power generation methods, are gaining attention. However, to ensure the practical operational efficiency and economic viability of these modular reactors, a comprehensive layout strategy considering various factors such as geographical conditions, demand patterns, power grid connectivity, and maintenance accessibility is necessary. This study models this problem from a multi-criteria decision-making and mathematical optimization perspective and proposes an optimal deployment model based on mixed integer linear programming (MILP). This study provides practical decision-making support tools for energy supply planning, power infrastructure expansion, and nuclear energy policy formulation, and can serve as foundational data for integrating future smart grids and renewable energy systems.

3 - Optimal Resource Allocation for Hyperscale Data Centers in Texas: a Deterministic Milp Approach

Ryan Callahan, University of California, Davis, Davis, CA, United States, Jiadong Zhao, Thanakorn Piyayothin, Aaron Rabinowitz

We develop a deterministic Mixed-Integer Linear Programming (MILP) model to determine the optimal energy infrastructure investment for projected hyperscale data centers while minimizing costs. The model combines long-term capital investment and short-term operational decisions with hourly dispatch dynamics for an entire year (8,760 time steps). Decision variables include binary capacity selection, continuous generation dispatch, and market participation in real-time and day-ahead wholesale energy markets. The model objective is to reduce total system costs, defined by a Levelized Cost of Energy (LCOE) framework, that includes fixed capital costs, operating and maintenance expenses, market energy prices, and fuel inputs.

Assuming collocation of these generation sources, our findings indicate that the optimal portfolio includes 54.4% natural gas, 36.1% solar, wind 9.5%, and no capacity for nuclear. The model installs 43.9 GW of on-site capacity but fails to meet the full 62.8 TWh demand and necessitates a grid import of 8.7 TWh. The annualized LCOE is \$0.8674/kWh, which is about 8x higher than regional wholesale market rates. Total infrastructure investment exceeds \$1.5 trillion.

Even with optimal co-location strategies, our analysis shows that grid purchases from legacy assets, whose capital costs have already been amortized, using existing Locational Marginal Prices (LMPs) remain more cost-effective than new-build supply. This formulation offers a decision-support framework for assessing the viability of vertically integrated power systems in data centers. Additionally, our model suggests hyperscale data centers may not be economically viable without significant policy intervention, cost innovation, or externalized cost bearing.

4 - Optimization of Energy Transition System Based on Process Linkage Modeling

Hao Ding, Nanjing University of Aeronautics & Astronautics, Nanjing, China, People's Republic of, Dequn Zhou

With the increasing penetration of variable renewable energy in electricity system, multiple uncertain risk factors, ranging from internal technology costs to external policies and markets, introduce significant revenue risks. This work constructed an optimization model by introducing various uncertainties to reach an optimal transition pathway toward carbon neutrality in electricity system considering multi-risks. The results indicate that risks associated with VRE penetration will increase the transition cost, but decision-makers' risk averse will actually reduce the cost. Based on the above findings, we emphasize the importance of risk-averse modelling in power system transition planning and support the future achievement of ambitious high VRE penetration goals. We suggest further reduction in technology costs and deployment of diverse flexible options. Various technologies' utilization is important to deal with the risks in electricity transition toward carbon neutrality.

5 - Optimization of Distribution Grid Restoration with Job Scheduling, Resource Dispatching, and Transportation Network Disruptions

Brent Austgen, Sandia National Laboratories, Albuquerque, NM, United States, Ross Guttromson

The restoration of distribution grids following disruptions is a complex challenge that spans multiple decision domains, including grid operation, job scheduling, and resource dispatching. We present an optimization model that incorporates decisions and constraints from these domains. The grid is represented as a graph, with isolable sections as nodes and switches as edges, allowing for effective management of power flow and topology configurations. The resource dispatching model accounts for crew capabilities, work site accessibility, site-to-site travel times, and disruptions in the transportation network. The job scheduling model, which links resource dispatching and grid operation, incorporates resource and time requirements, job dependencies, and work site safety. The model's objective function is to maximize criticality-weighted load served and crew safety. We intend for the model's job scheduling and resource allocation solutions to inform restoration planning, enabling utilities to respond more effectively to disruptions and minimize downtime for customers. We demonstrate the model using a case study featuring a small distribution feeder, a road network, and a simulated wildfire contingency. We discuss key aspects of the model solution, how the model fits in a rolling horizon optimization framework, worthwhile modeling extensions, and value to utilities.

WA69

Building B Level 4 B409

Intelligent Decision-Making and Organizational Strategy

Invited Session

MSOM: Technology, Innovation, and Entrepreneurship

Chair: Kostas Bimpikis, Stanford GSB, Stanford, United States

Co-Chair: Alp Sungu, Wharton, Philadelphia, PA, United States

1 - Strategic Supply-Demand Mismatch Through Lead Time and Pricing Autonomy: Evidence from a Car-Sharing Platform

Asif Nazrul, University of Cambridge, Cambridge, United Kingdom, Stylianos Kavadias

This paper examines supply-demand mismatch on hybrid peer-to-peer (P2P) car-sharing platforms, where decentralized hosts make strategic decisions about asset availability. Using booking-level data from one of the largest online marketplaces and exploiting a quasi-experimental policy shift in February 2023 that granted pricing autonomy to hosts, this study investigates how self-pricing affects listing behavior and utilization. A Triple Difference-in-Differences (DDD) framework identifies heterogeneous treatment effects moderated by host experience. The results show that experienced hosts with self-pricing privileges strategically delay their listings, reducing lead times by an additional 0.92 days compared to other host groups. This temporal shift contributes to a 6.9 percentage point decline in utilization, reflecting lower booked hours relative to availability. Inexperienced self-pricing hosts also experience diminished performance, likely due to suboptimal pricing decisions. Together, these findings suggest that pricing autonomy encourages strategic but inefficient supply behavior, particularly among experienced providers. This study contributes to the literature on marketplace design and decentralized operations by offering causal evidence on the behavioral consequences of pricing flexibility. It highlights how autonomy can exacerbate service availability gaps, particularly in advance-booking settings where supply timing is a critical decision that is overlooked in current literature. The findings underscore the need for experience-sensitive interventions, such as feedback-based incentives or dynamic listing nudges, to maintain platform efficiency while preserving host flexibility. Future work can extend this analysis using unsupervised clustering to uncover latent experience profiles and improve behavioral targeting in two-sided digital markets.

2 - Managing Multi-Component Projects: Task Prioritization and Optimal Stopping

Sasa Zorc, University of Virginia, Charlottesville, VA, United States, Vladimir Smirnov, Ilia Tsetlin

This paper examines a class of multi-component projects in which all components need to be successfully completed in order to realize the reward. Each component evolves stochastically over time and carries the risk of failure. We model this as a dynamic optimal control problem with endogenous decisions over how to distribute effort between components and when to abandon the project. We derive optimal component prioritization and stopping strategies. Since some of the optimal prioritization strategies are computationally demanding and with no clear managerial insight, we put additional focus on (a) specific cases where insight is clear (b) simple index policies that approximate the optimal ones, while being easy to interpret and compute. Our framework applies to a wide range of real-world problems, including research and development, software engineering, and academic research planning, where progress is uncertain, sequential, and interdependent.

3 - The Role of Discussions in Information-Gathering and Decision-Making

Panos Markou, UVA Darden School of Business, Charlottesville, VA, United States, Chan Tian, Hui Sun

We study how discussions among a panel of individuals aids in information gathering and making decisions. Leveraging verbatim transcripts from expert committees, we apply natural language processing techniques to analyze how discussion patterns inform influence and judgment.

4 - The AI Will See You Now: How Healthcare's Digital Revolution Rewrites Power

Emisa Nategh, Georgetown University, Washington, DC, United States

This paper reveals how healthcare AI's boom-bust cycles are not mere technological transitions but profound socio technical dramas that redistribute power between tech giants, healthcare systems, patients, and democratic institutions. Through the lens of "agency loops"—five-phase recursive patterns where humans and AI continuously renegotiate control—we expose how healthcare organizations navigate an impossible trilemma: achieving AI-mandated Medicare compliance, preserving patient agency, and maintaining democratic accountability. Our research unmasks the hidden choreography behind healthcare AI adoption, moving beyond deterministic views of technology to highlight how algorithms are shaped by, and in turn, shape everyday work practices. Each innovation "spring" creates unstable sociomaterial assemblages dominated by tech oligopolies (Google Health, Epic, Microsoft) that control crucial "choke points" in the AI ecosystem. These cycles inevitably collapse into "winters" when attributing life-and-death decisions to opaque algorithms proves ethically unsustainable.

Wednesday, October 29, 9:30 AM - 10:45 AM

WB01

Building A Level 3 A301

Data Strategy Professionals/Timefold

Invited Session

Technology Showcase

Chair: Loraine Coleman, INFORMS, 5521 Research Park Drive, Catonsville, MD, 21228, United States

1 - Business Value of AI Governance Frameworks

Nicole Janeway Bills, Data Strategy Professionals, Alpharetta, GA, United States

The talk addresses critical knowledge gaps that many data practitioners face in navigating AI Governance. The presentation will provide insights into the strengths and weaknesses of various AI Governance frameworks from organizations such as NIST and OECD and companies such as Microsoft, AWS, and Google. Moreover, we'll discuss practical recommendations for the application of these frameworks to mitigate risk and enhance business value.

2 - VRP: When Research Meets Reality

Geoffrey De Smet, Timefold BV, Gent, Belgium

It's one thing to solve a VRPTW (Vehicle Routing Problem with Time Windows). It's another to put it in production for a fleet of 50,000 technicians and save hundreds of millions of dollars per year. It's not just more constraints/objectives and more data. In this session, we will talk about the business and technical challenges of transforming an academic VRP into a solution that works well in production.

WB02

Building A Level 3 A302

Resilience and Disruption Management in Transportation Systems

Contributed Session

Contributed

Chair: Nazli Demir, University of Waterloo, Waterloo, ON, Canada

1 - Travel Behavior Modeling Under Network Disruption: a Computational Graph Approach

JINGYAN LI, Carnegie Mellon University, Pittsburgh, PA, United States, Sean Qian

Large-scale disruptions have significant and lasting impacts on transportation systems. Understanding the day-to-day evolution of traffic patterns from the perspective of individual traveler behavior provides essential insights into how traffic systems respond to disruptions over time. Although recent research has explored day-to-day traffic modeling under network disruptions, a comprehensive, end-to-end calibration using real-world data remains largely unexplored.

In this study, we address this gap by proposing a computational graph-based approach to model travelers' day-to-day behavior under network disruptions. Specifically, we construct an integrated framework comprising a travel behavior model, a route choice model, and a dynamic network loading component. This framework captures daily

traffic dynamics as a discrete-time Markov chain process.

Our approach offers notable advantages compared to existing studies. Firstly, the computational graph methodology enables an end-to-end calibration of the entire traffic assignment model, directly learning traveler disutility functions from observed data. Secondly, our design of disutility functions emphasizes generalizability, allowing the model to effectively predict traffic responses to out-of-sample disruptions within the same transportation network.

We have successfully trained and validated the proposed model using a synthetic network, where the model empirically shows the capability of returning to the equilibrium. Future work will involve applying our methodology to real-world data sets, thereby validating its practical effectiveness and predictive capability.

The outcomes of this research hold significant potential for enhancing resilience planning and decision-making processes in transportation management.

2 - Understanding Day-to-Day Evolution of Traffic Patterns During Disruption: An Incremental Non-Negative Matrix Factorization Approach

Jingyan Li, Canegie Mellon University, Pittsburgh, PA, United States, Sean Qian

Recent major infrastructure failures have resulted in significant socio-economic losses and sudden disruptions to transportation networks. Traffic patterns are altered profoundly after the disruption, ultimately reaching a new equilibrium as travelers adjust to the changed infrastructure. However, accurately modeling this disruption-to-equilibrium transition is challenging due to concurrent disturbances from commuting peaks, weather conditions, special events, and accidents. Removing these noise factors, particularly from high-resolution spatiotemporal data, is critical for understanding true disruption impacts.

To address this challenge, this study proposes an incremental non-negative matrix factorization (NMF) approach to capture daily spatiotemporal traffic features during disruption periods. Unlike traditional static NMF methods, our approach dynamically preserves critical features from previous days while simultaneously incorporating emergent patterns resulting from disruptions. This incremental learning and memory-refreshing mechanism facilitates an interpretable understanding of how traffic patterns evolve over time.

The method is validated with data from the 2023 I-95 highway collapse in Philadelphia, Pennsylvania. Results indicate that equilibrium was restored within one week following each disruption event. During the recovery, new features reflecting increased congestion around closure points and along alternative routes emerged, while pre-disruption bottleneck features gained greater significance. Conversely, previously dominant features indicating lighter traffic conditions along alternative routes diminished or disappeared altogether.

This incremental NMF method enables real-time adaptation immediately following disruptions, making it a valuable tool for proactive traffic management. Furthermore, the interpretable low-rank features identified by this method can serve as representative scenarios for conducting what-if analyses, enhancing preparedness, and minimizing adverse impacts in future disruption events.

3 - Adaptive Framework for Flood Relief Using Drones and Floating Mini Depots for Equitable Aid Distribution

Fateme Hafizi, Illinois Institute of Technology, Chicago, IL, United States, Mohammad Miralinaghi

Floods pose significant logistical challenges in delivering essential goods such as food, water, and medicine to affected populations, especially in underserved and isolated areas where transportation infrastructure is damaged or completely submerged. This study presents an innovative disaster relief strategy employing drones and floating mini depots to ensure equitable and efficient distribution of essential supplies during flooding disasters. Mini depots, deployed on boats, serve as mobile hubs that store and dispatch emergency goods to affected communities. Drones launched from mobile depots can rapidly reach affected regions, reducing aid delivery time and improving the overall system's responsiveness. By removing dependence on road-based infrastructure, this system enables access to areas that would otherwise remain isolated during severe flooding. Historical flood and population data guide the identification of high-demand regions, enabling strategic preparedness to minimize delays during disaster response. Predictive analytics and machine learning are employed to anticipate demand surges and optimize routing, ensuring minimal delays and effective resource utilization. This strategy enhances the effectiveness of emergency logistics during flood events while promoting equity by prioritizing access for marginalized populations. This strategy improves the capacity to maintain supply chains during disruptions to conventional delivery routes, making it particularly advantageous for regions susceptible to recurrent flooding. The integration of drone technology with adaptable, aquatic distribution centers constitutes an approach to addressing significant logistical challenges encountered in disaster relief efforts. The system provides a flexible framework that can be adapted to diverse geographical contexts and enhances resilient and participatory emergency response strategies during climate-related disasters.

4 - Navigating Network Availability: a Shortest Path Approach to Uav Traffic Routing

Nazli Demir, University of Waterloo, Waterloo, ON, Canada, Fatma Gzara

The use of unmanned aerial vehicles (UAVs), commonly known as drones, is rapidly increasing across both commercial and recreational domains. This growth highlights a significant challenge: the lack of a drone traffic management system with a well-defined airspace traffic network and, routing and flight path planning. As drone operations expand, airspace becomes increasingly crowded, increasing the risk of aerial conflicts and emphasizing the need for effective regulatory and operational frameworks. This research aims to address these challenges by developing tools to be integrated into UAV Traffic Management (UTM) systems, enabling efficient flight planning for UAVs within an airspace structured as a network in consideration of ongoing traffic. A novel mathematical model is proposed to find a flight plan that specifies the flight path as well as a time schedule of traversing arcs on the path. A network transformation method is developed to reformulate the problem on an auxiliary network that allows solution by a novel labeling algorithm. Computational experiments demonstrate that our proposed solution approach outperforms a commercial solver by efficiently solving large-scale problems in under one second, making it highly suitable for seamless integration into UTM systems. Our findings offer valuable insights into determining the optimal flight paths of UAVs across various problem settings, such as the distance between origin and destination, the duration of the operation horizon, and network traffic. These insights may guide UTM providers in designing UTM systems that facilitate efficient airspace management.

WB03

Building A Level 3 A303

Emergency Response

Contributed Session

Contributed

Chair: Shuai Shao, University of South Florida, Tampa, FL, United States

1 - Stochastic Optimization of Hurricane Relief Logistics with Scenario-Based Planning and Rolling Horizon Updates

Ali Tolooie, Mercer University, Macon, GA, United States, Marzieh Soltanolkottabi

This study proposes a stochastic optimization framework for managing FEMA's multi-tiered humanitarian logistics network in response to hurricanes. The model integrates strategic preparedness and operational responsiveness using a rolling-horizon, two-stage stochastic programming approach. In the preparedness phase, optimal inventory pre-positioning at Major Distribution Centers (MDCs) is determined, and candidate locations for Pre-Staging Areas (PSAs) are identified prior to hurricane season. Once a hurricane forms, operational decisions are revised every six hours based on updated forecasts of the storm's trajectory, intensity, and potential impact.

Scenario generation is based on historical hurricane data and ensemble forecasting, with spatial impact assessments derived using FEMA's Hazus-MH platform. These scenarios incorporate uncertainties in infrastructure accessibility, facility operability, and time-dependent regional demand. The model also addresses perishability of items, penalties for unmet demand, return logistics, and disposal of expired supplies, while minimizing transportation and holding costs.

By dynamically integrating real-time data into decision-making, the proposed framework supports more responsive and cost-effective disaster logistics. It offers a practical tool to enhance FEMA's ability to deliver critical supplies under uncertainty, improving both preparedness and post-disaster outcomes.

2 - A Data-Informed Optimization Approach for Dynamic Fire Unit Relocation

Michael Vieira, California Polytechnic State University, San Luis Obispo, Livermore, CA, United States, Zhiyuan Wei

Fire departments face critical challenges when responding to major incidents that require mass deployment of resources, leaving large areas with insufficient coverage for subsequent incidents. Traditional strategies are usually reactive, responding to events that have already occurred. As a result, they overlook future incidents that could occur in the service area of unavailable stations, leading to delayed response and uncovered demand. A common way to address this issue is through a proactive strategy that temporally redistributes units during resource-intensive events. However, this proactive strategy is often based on the prior experience of dispatchers, which is subjective. To address this research gap, we aim to develop a data-driven approach for dynamically relocating fire units to ensure efficient coverage. Specifically, we propose an optimization model that determines optimal relocations to improve coverage. The methodology integrates big data on fire unit availability with geographic coverage requirements to maximize weighted coverage of demand nodes while respecting operational constraints. Our model accounts for various factors including station capacity, travel times between stations, and high-risk areas based on expected demand patterns. The performance of relocation strategies is evaluated in terms of both efficiency and equity. To enable fire departments to maintain the most effective coverage possible, the proposed data-driven optimization framework reallocates emergency service resources under uncertainty to enhance operational resilience and public safety.

3 - Optimal Initial Dispatch of Aerial Firefighting Resources in the Republic of Korea Under Wildfire Spread Simulation

Chanjung Lee, Seoul National University, Seoul, Korea, Republic of, Seung-Jae Joo, Junyeong Choi, Yohan Lee

Wildfires in the Republic of Korea have intensified in frequency, scale, and simultaneity, creating complex challenges for rapid aerial response. We propose an integrated framework that combines (i) an integer-linear optimization for the initial dispatch of helicopters—minimizing immediate suppression and damage costs while enforcing travel-time, refueling, and “golden-hour” response constraints—with (ii) a mixed-integer module that refines fleet composition to reduce the present value of operational and damage expenses over a planning horizon. While prior research typically treats dispatch and fleet sizing separately, our unified formulation captures their interdependencies under realistic operational constraints. To capture fire dynamics, we integrate a quasi-physical wildfire-spread model calibrated on nationwide historical fire data (2022–2024). Numerical experiments demonstrate that our approach improves average response times and reduces expected burned area compared with heuristic rules, all within computation times suited for operational use. Our results also illustrate how relaxing night-flight limits or adding heavy-lift helicopters influences both short-term effectiveness and long-term cost efficiency. Designed for Korea's operational context, our framework delivers a scalable decision-support tool capable of interfacing with advanced wildfire-spread models and incorporating state-of-the-art approaches such as reinforcement learning. Its modular design also supports future extensions to coordinate ground resources, enabling more integrated emergency-response planning.

4 - Designing Resilient Firefighting Systems: Strategic and Tactical Planning for Wildfire Suppression Using a Tri-Level Optimization Framework

Payam Khazaelpour, Florida International University, Maimi, FL, United States, Shabnam Rezapour Behnagh, Ania San Roman

In the face of escalating wildfire threats driven by climate change and human encroachment into fire-prone regions, strategic planning for firefighting has become a critical component of disaster preparedness and response. This study presents a comprehensive tri-level optimization framework to address the complex challenges of resource allocation and operations in wildfire suppression. The proposed model integrates three interdependent decision-making layers: (1) a pre-disaster planning stage, in which a regional authority (the “defender”) determines the optimal number and locations of fire stations and water reservoirs; (2) a worst-case scenario simulation, where a hypothetical “attacker” models the spread of wildfire across zones over a time horizon to challenge the defender's decisions by maximizing disruption; and (3) a post-disaster response phase, where tactical decisions are made regarding the dispatch and routing of firefighting vehicles to extinguish the fire quickly and efficiently. The outer level minimizes total response costs and fire suppression time, balancing economic efficiency with rapid containment. The middle level simulates an adversarial fire pattern to capture risk-averse planning under uncertainty. The inner level coordinates vehicle movements and firefighting actions to reduce logistical costs and accelerate fire termination during emergency response.

By bridging infrastructure planning with dynamic emergency operations, this framework offers a robust and adaptable approach to wildfire preparedness. It provides valuable insights for policymakers, emergency managers, and urban planners engaged in regional disaster risk management.

5 - Commit-OR-Defer Shortest Path Interdiction

Shuai Shao, University of South Florida, Tampa, FL, United States, Juan Borrero, Oleg Prokopyev

We study a sequential shortest path interdiction problem where an interdicator, with initially uncertain knowledge of arc costs, aims to maximize an evader's cumulative travel costs over several time periods. The evader possesses complete cost information. The interdicator has a limited budget for permanent arc removals and refines her cost estimates by observing the evader's path choices. We analyze different strategies for resource deployment.

Our analysis reveals that when adopting worst-case or best-case cost outlooks, a fully-committing strategy is optimal for the interdicator. When performance is evaluated across multiple potential cost scenarios, a fully committing strategy is optimal if considering up to two such scenarios; however, for a larger and more diverse set of potential costs, a fully deferring approach yields better results. We further introduce and analyze ‘risk-free’ policies, which aim to guarantee a specific performance threshold against any realization of uncertain costs, and develop mixed-integer programming formulations for these policies over short time horizons. Computational experiments evaluate the performance of 1-time-step and 2-time-step risk-free strategies, alongside fully-committing (optimistic and pessimistic) and fully-deferring policies, highlighting their practical tractability and comparative effectiveness.

WB04

Building A Level 3 A304

Risk Management and Decision Models

Contributed Session

Contributed

Chair: Kenneth Bruninx, TU Delft, Delft, Netherlands

1 - Combining Benders Decomposition and Stochastic Dual Dynamic Programming: Novel Approaches for Enhanced Performance

Anil Kaya, Karlsruhe Institute of Technology, Karlsruhe, Germany, Alexander Shapiro, Steffen Rebennack

We propose novel algorithmic ideas for combining Benders decomposition and Stochastic Dual Dynamic Programming (SDDP). These algorithms are suitable to solve multi-stage stochastic mixed-integer programming problems of special structure. Here, the first-stage decisions are mixed-integer while other stages are linear programming problems (e.g. power expansion planning models). Benders decomposition is an efficient solution technique for two-stage optimization problems with a block-angular structure, where the so-called subproblems are linear programming problems. SDDP can efficiently solve multi-stage stochastic linear programming problems. Thus, combining these two methods leverages the two-stage nature of the problem as well as its multi-stage structure. However, when combining Benders decomposition with SDDP, the optimality cut for the master problem is not guaranteed to be tight or valid. To this end, we present novel algorithmic ideas to obtain exact optimality cuts. Then, we apply the proposed approaches to power expansion planning problem (as part of the ARTESIS project) to compare their performance and demonstrate their efficacy in addressing large-scale multi-stage stochastic mixed-integer problems.

2 - Game-Theoretic Analysis of Risk-Averse Behaviors in Power Battery Recycling: Closed-Loop Supply Chains with Government Subsidies

Dengke Bai, Ningbo University of Technology, Ningbo, China, People's Republic of, Li Hou

Abstract: The rapid growth of electric vehicles has led to a surge in end-of-life power batteries, posing significant environmental and resource challenges. Efficient recycling is critical not only for sustainable development but also for securing strategic raw materials like lithium and cobalt. While government subsidies are widely used

to promote battery recycling, their interaction with the risk-averse behaviors of supply chain members remains poorly understood. This study develops a tripartite Stackelberg game model incorporating stochastic demand and cost uncertainty, with the manufacturer as the leader and the retailer as the follower. The model examines how risk aversion and subsidies interact through nonlinear pricing dynamics under stochastic demand. We analyze four policy scenarios combining manufacturer and retailer risk aversion with subsidized and unsubsidized conditions. Using mean-variance analysis and backward induction, we derive equilibrium pricing strategies and quantify their impacts on recycling efficiency and supply chain performance. Key findings reveal that (i) government subsidies effectively mitigate risk-averse behaviors while optimizing pricing strategies, with subsidy effectiveness contingent on market uncertainty levels; (ii) risk aversion creates a dual effect on pricing—increasing prices in high-cost scenarios while promoting market expansion through lower prices when production costs are favorable; (iii) consumer recycling sensitivity and voluntary return rates nonlinearly influence optimal subsidy design, with subsidies showing greater impact in high-uncertainty markets. Based on these insights, we propose a tiered subsidy allocation framework: (i) higher subsidies to recyclers in early stages to break risk cascading effects, (ii) manufacturer subsidies tied to recycling rate thresholds.

3 - Modeling Robust Energy Systems Considering Weather Uncertainty and Nuclear Power Failures: a Case Study in Northern Europe

Lina Reichenberg, Chalmers University, Göteborg, Sweden, Kamran Forghani, Xiaoming Kan, Fredrik Hedenus

Capacity expansion models used for policy support have increasingly represented both the variability and uncertainty of weather-dependent generation (wind and solar). However, although also uncertain, as demonstrated by the performance of the French nuclear power fleet in 2022, uncertainty arising from nuclear power outages has been largely neglected in the literature. This paper presents the first capacity expansion model that considers uncertainty in nuclear power availability caused by unplanned outages. We propose a mathematical model that combines a scenario-based stochastic optimization approach (to deal with weather-related uncertainties) with a data-driven adjustable robust optimization approach (to deal with nuclear failure-related uncertainties). The robust model represents the bulky behavior of nuclear power plants, with large (1 GW) units that are either on or off, while at the same time letting the model decide on the optimal amount of nuclear capacity. We tested the model in a case for Northern Europe (seven nodes) with a time resolution of 1250 time steps. Our findings show that nuclear power outages do, in fact, impose a vulnerability on the energy system if not considered in the planning phase. Our proposed model performs well and finds solutions that prevent Loss-of-Load (at a price of robustness of 0.6%), even in more extreme weather conditions. Robust solutions are characterized by a higher capacity of gas plants, but, perhaps surprisingly, nuclear power capacity is barely affected.

4 - Transmission Expansion Planning Under Wildfire Risk Uncertainty: A Two-Stage Stochastic Optimization Approach

Easha Rajalaxmi Vivesh, The George Washington University, Washington DC, DC, United States, Ekundayo Shittu, Miguel Lejeune

This research presents a novel approach to transmission expansion planning (TEP) that incorporates wildfire risk uncertainty using a two-stage stochastic optimization framework. As climate change increases wildfire frequency and severity, power utilities face growing challenges in maintaining grid reliability while minimizing fire risks. The proposed methodology integrates probabilistic wildfire modeling with TEP to determine optimal investment strategies balancing costs against resilience. The optimization model is a two stage stochastic model: first-stage "here-and-now" transmission line investment decisions made before wildfire scenarios are realized, and second-stage "wait-and-see" operational decisions (generation dispatch, load shedding, etc.) implemented after wildfire conditions materialize. The objective function minimizes investment costs plus expected operational costs across all scenarios, subject to power flow physics and risk-based constraints. Using Monte Carlo simulation to generate realistic wildfire scenarios based on temperature, humidity, and wind speed distributions, the model captures the complex relationship between environmental conditions and wildfire probability. A case study using the Garver 6-bus test system demonstrates the model's applicability, with sensitivity analysis exploring the impact of varying risk thresholds on investment decisions, operational costs, and load shedding. Results indicate a significant trade-off between wildfire risk tolerance and system costs, with optimal solutions found in the 0.4–0.6 risk threshold range. This work provides power system planners with a quantitative framework to develop transmission networks resilient to increasing wildfire threats.

5 - Penalty-Free SDDP: Incorporating Benders Feasibility Cuts for Robust Multi-Stage Operational Planning

Guilherme Freitas, PSR, Rio de Janeiro, Brazil, Alexandre Street, Luiz Carlos Costa Junior, Tiago Andrade

The increasing complexity of energy systems stems from the global energy transition, integration of intermittent renewable sources, and the need to ensure supply security and reliability. In Brazil, with its predominantly hydroelectric matrix, incorporating large-scale solar and wind sources presents additional challenges.

Brazilian electrical system operation requires constant balance between economic efficiency and meeting various constraints such as water resource management, environmental preservation, and supply security. This conflict manifests in discussions between different regulatory agencies with distinct and often difficult-to-reconcile objectives. There is a growing need for methodologies that bring rationality to decision-making processes, enabling adequate quantification of each restriction's impact.

The Stochastic Dual Dynamic Programming (SDDP) algorithm has established itself as the primary tool for multi-stage operational planning under uncertainty. The traditional approach involves introducing penalties for constraint feasibility, but this presents significant challenges. Penalty calibration is subjective since there is often no directly associated economic value, making definitions arbitrary and potentially distorting economic signals and algorithm convergence.

This work proposes an SDDP formulation enhancement without artificial penalties, adding Benders feasibility cuts within the algorithm's recursive process. Analogous to the Future Cost Function, the approach introduces a Future Feasibility Function, indicating the benefit of anticipating storage for future constraint compliance.

Results from the Brazilian case study demonstrate significant improvements over the traditional approach, eliminating artificial penalties while improving marginal cost precision and convergence, ensuring optimization based on operational costs aligned with physical and economic constraints. The approach successfully maintains hydraulic constraint compliance while providing more realistic operational decisions.

6 - AI-Augmented Monte Carlo Simulation for Capital Investment Decisions in Innovation Project Portfolios

Gustavo Vinueza, Ferryfield Group, London, United Kingdom

Capital investment decisions in innovation project portfolios are characterized by high uncertainty—ranging from technical feasibility and development timelines to market adoption and regulatory approval. Traditional deterministic models often fail to capture these uncertainties, leading to suboptimal capital allocation and increased risk exposure. This paper introduces a decision-support framework that combines Monte Carlo Simulation (MCS) with artificial intelligence (AI) to support more robust, data-informed investment strategies.

Monte Carlo Simulation is used to model uncertainty across financial and scheduling dimensions, producing probabilistic estimates for key metrics such as Net Present Value (NPV), Internal Rate of Return (IRR), and the likelihood of cost and time overruns. To improve simulation fidelity, AI techniques—particularly machine learning—are applied to historical and contextual project data. These models help refine input distributions, detect interdependencies between projects, and dynamically adjust assumptions in response to new evidence.

Beyond input calibration, AI is used to support optimization efforts by analyzing simulated outcomes and identifying portfolio strategies that maximize expected value under defined risk constraints. This enables decision-makers to explore trade-offs between risk and return, better allocate limited capital, and gain visibility into the impact of uncertainty drivers.

The proposed framework is demonstrated through a fictitious innovation portfolio composed of diverse early-stage projects. Results show improved transparency, better risk-adjusted returns, and actionable insights for capital allocation under uncertainty. The methodology is applicable to industries where innovation drives value, such as pharmaceuticals, energy, and technology.

7 - Risk-Aversion and Capacity Remuneration Mechanisms in Electricity-Hydrogen Markets

Kenneth Bruninx, TU Delft, Delft, Netherlands, Alessio Berdin, Laurens de Vries, Aad Correljé

Hydrogen and derived fuels may act as long-term energy storage in climate-neutral energy systems. However, risk-averse investors will not invest in sufficient renewable electricity, back-up, electrolyzer and storage capacity if they are only remunerated for the hydrogen or electricity produced. We develop a stochastic equilibrium model to study whether capacity markets can restore investments to their risk-neutral levels. Our results show that the efficacy of capacity markets depends on complementary instruments to ensure the availability of renewables. If risk-averse investors build less renewables than what is societally optimal, capacity markets in the electricity and hydrogen sectors are needed to restore the overall capacity mix and limit consumer costs. If complementary instruments lift investments in renewables to their societally optimal level, a capacity market in the electricity sector suffices. In this situation, an additional capacity market in the hydrogen sector may trigger a bias toward hydrogen-fired backup capacity. This illustrates that an integrated systems' perspective is required to design future energy markets.

WB05

Building A Level 3 A305

Algorithm Development

Contributed Session

Contributed

Chair: Tianjian Huang, University of Southern California, Los Angeles, CA, United States

1 - A Scalable Deterministic Global Optimization Algorithm for Constraint Clustering Problems

Pedro Chumtipatz, University of South Florida, Tampa, FL, United States, My Duong, Cristobal Heredia, Kaixun Hua

Constrained clustering leverages limited domain knowledge to improve clustering performance and interpretability, but incorporating pairwise \emph{must-link} and \emph{cannot-link} constraints is an NP-hard challenge, making global optimization intractable. Existing mixed-integer optimization methods are confined to small-scale datasets, limiting their utility. We propose Sample-Driven Constrained Group-Based Branch-and-Bound (SDC-GBB), a decomposable branch-and-bound (BB) framework that collapses must-linked samples into centroid-based pseudo-samples and prunes cannot-link through geometric rules, while preserving convergence and guaranteeing global optimality. By integrating grouped-sample Lagrangian decomposition and geometric elimination rules for efficient lower and upper bounds, the algorithm attains scalability via embarrassingly simple parallelism. Experimental results show that our approach handles datasets with 200,000 samples for cannot-link constraints and 1,500,000 samples for must-link constraints, which is 200 - 1500 times larger than the current state-of-the-art under comparable constraint settings, while reaching an optimality gap of < 3%. In providing deterministic global guarantees, our method also avoids the search failures that off-the-shelf heuristics often encounter on large datasets.

2 - QOCO: A Quadratic Objective Conic Optimizer with Custom Solver Generation

Govind Chari, University of Washington, Seattle, WA, United States, Behcet Acikmese

Second-order cone programs (SOCPs) with quadratic objective functions are common in optimal control and other fields. Most SOCP solvers which use interior-point methods are designed for linear objectives and convert quadratic objectives into linear ones via slack variables and extra constraints, despite the computational advantages of handling quadratic objectives directly. In applications like model-predictive control and online trajectory optimization, these SOCPs have known sparsity structures and require rapid solutions. When solving these problems, most solvers use sparse linear algebra routines, which introduce computational overhead and hinder performance. In contrast, custom linear algebra routines can exploit the known sparsity structure of problem data and be significantly faster.

This work makes two key contributions: (1) the development of QOCO, an open-source C-based solver for quadratic objective SOCPs, and (2) the introduction of QOCOGEN, an open-source custom solver generator for quadratic objective SOCPs, which generates a solver written in C that leverages custom linear algebra. Both implement a primal-dual interior-point method with Mehrotra's predictor-corrector. Our benchmarks show that QOCO is faster and more robust than many commonly used solvers, and solvers generated by QOCOGEN are significantly faster than QOCO and are free of dynamic memory allocation making them an attractive option for real-time optimization on resource-constrained embedded systems. Both QOCO and QOCOGEN are easy to use and can be called from the modeling language CVXPY.

3 - Modified Eigenvalue Method for Nonconvex Miqucqp and Parallel Local Branching

Yongzheng Dai, The Ohio State University, Columbus, OH, United States

The ability to quickly find feasible solutions is an essential component of modern MIP solvers. A good feasible solution early in the solution process can significantly reduce the size of the branch-and-bound tree. Moreover, for many end users — especially when solving hard problems such as nonconvex MIQCQPs — a good feasible solution can be more valuable than a tight dual bound. Known primal heuristics for MIPs are typically based on projection onto the relaxed feasible regions or local search on reduced problems, which often get into trouble with nonconvex MIQCQPs because of nonconvex constraints. In this work, we propose a modified eigenvalue-based method to approximate nonconvex MIQCQPs, along with two enhanced feasibility pump algorithms designed to rapidly identify feasible solutions. To further improve the quality of these solutions, we develop a parallel local branching approach that efficiently explores large neighborhoods. Computational experiments demonstrate the effectiveness and promise of the proposed methods.

4 - Genlars: An Efficient Path-Seeking Algorithm for Parameterized Quadratic Programs

Roufeng Liu, MIT, Cambridge, MA, United States, Haihao Lu, Rahul Mazumder

We introduce a path-seeking algorithm for efficiently solving parameterized quadratic programs (QPs) with varying parameters where the parameter could appear in regularization terms, constraint right-hand-sides, or linear terms in objectives. Building on the piecewise linearity of solutions in such problems—generalized from LASSO-like settings to constrained QPs with equality and bound constraints—our approach leverages the KKT conditions and dynamic partitioning of variables into different active sets. By tracing linear segments between critical events, the algorithm efficiently updates solutions while maintaining sparsity and scalability, particularly in high-dimensional regimes where the number of constraints and rank of the quadratic term in the objective function is far fewer than the number of decision variables. We also propose a novel gradient screening technique for screening out inactive variables, which provides additional speedups for sparse solutions. In addition, we propose a way to efficiently update the factorization of the coefficient matrix of the KKT system as the active set changes, which allows the system to be solved in quadratic time. Benchmarks demonstrate significant speedups over Gurobi in sparse settings, including box-constrained LASSO and Markowitz portfolio optimization, and SVMs, highlighting its utility in scenarios requiring rapid path exploration or iterative parameter tuning.

5 - A Graph-Based Approach for Representing, Comparing, and Classifying Integer Programs

Tara Schroth, University at Buffalo, Buffalo, NY, United States, Kyle Hunt, Chase Murray

It has long been observed that model structure may play an important role in determining the effectiveness of certain solution techniques to mixed integer linear programs (MILPs). One implication is that solution techniques that work well for one class of MILP are often found to be effective on other “similar” problems. Despite the potential benefits, there is limited research in the area of recognizing similarity among MILPs. This paper seeks to address this gap by presenting a framework for classifying and comparing instances of MILPs, based on their mathematical structure. A new graph-based representation of MILPs is proposed, where decision variables and constraints are described by nodes, and arcs denote the presence of decision variables in certain constraints. Inspired by recent advances in the area of deep learning for graph-structured data, two methods of leveraging the MILP graph representations for classification and comparison are provided. In the first relational approach, a graph convolutional network (GCN) is employed to classify these so-called MILP graphs as having come from one of a known number of problem types. The second approach makes use of latent features learned by the GCN to compare graphs to one another directly. As part of the latter approach, we introduce a formal measure of graph-based structural similarity. Empirical studies indicate that both the classification and comparison procedures perform well on a test set. Additional properties of MILP graphs are also explored through a series of case studies.

6 - On Measuring and Predicting Performance Variability in Mixed Integer Programming (Mip) Solvers

Mourchid Adegbindin, Cornell University, Ithaca, NY, United States, Brenda Dietrich, Jamol Pender

Mixed-Integer Programming (MIP) solvers are widely applied to solve complex optimization problems across a broad range of domains. While these solvers are generally expected to behave consistently, they often exhibit surprising variability. Even under a priori neutral perturbations, a solver may complete the same problem in seconds during one run, yet take significantly longer—or fail to return a high-quality solution—in another.

Although solver variability has been known for decades, systematic investigation remains limited. In this work, we introduce the Weighted Variability Score (WVS), a unified and robust metric that quantifies solver instability across four dimensions: runtime, optimality gap, LP iterations, and branch-and-bound nodes. Unlike traditional runtime-based measures, WVS accounts for partial solutions, search effort, and remains robust to outliers.

We then conduct a large-scale empirical analysis on over 1000 MIPLIB instances. This experiment reveals that variability is neither rare nor restricted to hard instances; in fact, some known “easy” instances behave erratically, while others remain stable despite their complexity. Finally, we build a predictive modeling pipeline that demonstrates that a curated set of static and dynamic features can accurately predict WVS on most instances.

These results underscore the importance of treating solver variability as a first-class property in both algorithmic research and industrial deployment. By revealing that variability is not only systematic but also partially predictable, our findings offer a useful tool for preemptive tuning and solver configuration, lay a foundation for fairer evaluation of new algorithms, and support trust in resource-constrained applications where dependable behavior is essential.

7 - Fedsam: Sharpness-Aware Minimization for Improved Generalization Under FL Settings

Tianjian Haung, University of Southern California, Los Angeles, CA, United States, Maher Nouiehed

While generalization in Federated Learning (FL) has attracted growing attention in the machine learning community, it remains an emerging area of study. A key challenge arises from the inherent heterogeneity of client data and their varying computational capabilities. Recent research has drawn a connection between poor generalization and the sharpness of the optimization landscape.

In this paper, we propose a novel SAM-inspired framework tailored for FL. Unlike existing methods that apply SAM locally at each client, our approach penalizes the sharpness of the global loss function directly. We begin by illustrating a counter-example that highlights a critical limitation: flat minima at local clients do not necessarily aggregate into a flat global solution. This motivates our design of a sharpness-aware algorithm that adaptively estimates gradient similarity across clients and uses this to modulate the sharpness penalty. Specifically, each client receives a distinct penalty term based on the similarity of its gradients to those of other clients, accounting for data heterogeneity.

We provide theoretical analysis establishing the convergence of our proposed algorithm. To validate its effectiveness, we conduct extensive experiments on MNIST, FMNIST, and CIFAR datasets. The results demonstrate that our method achieves notable improvements in generalization performance compared to existing federated optimization techniques.

WB06

Building A Level 3 A306

Demand Forecasting and Sales Analytics

Contributed Session

Contributed

Chair: Prasun Singh, 95054

1 - Demand Forecasting Across Data Silos: Federated Shapley Valuation for Restaurant Supply Chain Collaboration

Yixuan Li, Johns Hopkins University, Baltimore, MD, United States

Demand fluctuations in restaurant supply chains are driven by heterogeneous and distributed data sources, including sales records, lead times, promotional activities, and weather conditions. These data are siloed across suppliers, central kitchens, and storefronts, giving rise to data islands, privacy concerns, and disputes over value attribution. This study proposes a collaborative modeling and data valuation framework that integrates Vertical Federated Learning with Shapley value theory, enabling cross-organizational demand forecasting without exposing raw data. Participating entities exchange only homomorphically encrypted gradients, ensuring commercial confidentiality. A Monte Carlo approximation of the Shapley value is embedded within each training round to quantify the marginal contribution of each party's feature set to model performance. This approach reduces the computational burden of traditional Shapley evaluation from exponential to linear complexity while preserving theoretical unbiasedness. The framework operates under an honest-but-curious threat model and supports game-theoretic mechanisms for revenue sharing based on information value. By unifying privacy-preserving distributed optimization with cooperative game theory, the proposed method offers both a theoretical advance and a practical solution for incentivizing data collaboration in multi-agent supply chains.

2 - Applied Econometric and Machine Learning Approaches for Sales Forecasting: A Case Study from A European Drugstore Chain

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

We analyze sales and related data from a European drugstore chain to develop predictive models for retail performance. After performing extensive feature engineering, we apply a range of econometric and machine learning techniques—including regression analysis, time series models, random forest, and XGBoost—to forecast sales. Our results demonstrate the effectiveness of combining structured data processing with advanced predictive analytics. The study offers a practical and adaptable framework that integrates econometric and machine learning approaches, contributing to data-driven sales and revenue management in retail and related industries.

3 - When Learning Less is More: Embedding Inventory Theories in End-to-End Deep Learning for Perishable Inventory Systems with Random Lead Time

Xuan Liao, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Zhenkang Peng, Ying Rong

Deep learning has shown promise in operations management field. In this paper, we show that training black-box deep neural networks to learn complex inventory state-to-action policies can be ineffective. This paper studies a perishable inventory system with random lead time and unknown demand and lead time processes. We introduce two end-to-end DNN-based policies—E2E-OQ, which directly outputs an order quantity, and E2E-PIL, which predicts a projected inventory level (the target level at the arrival point). Unlike E2E-OQ, the E2E-PIL network embeds a classical “lift-to-level” heuristic for projected inventory, constraining how the current inventory state influences decisions. Through extensive simulations, E2E-PIL consistently outperforms both traditional prediction-then-optimization benchmarks and the unconstrained E2E-OQ policy. We show that embedding a well-specified (though suboptimal) heuristic reduces sample complexity and stabilizes learning, with only marginal loss in approximation capability. This demonstrates the principle that, for complex management problems, “learning less is more”: incorporating domain knowledge into DNN architectures can yield more efficient and robust policies.

4 - Human-Machine Demand Forecasting

Thomas Kretschmar, WHU - Otto Beisheim School of Management, Vallendar, Germany, Arnd Huchzermeier

Human forecasters and forecasting algorithms perform well in different cases, especially if the human has additional information through private communication or domain knowledge. We propose a method to combine the two forecasts to leverage both of their strengths. A meta-learner selects either the human or the machine forecast based on time-series and product features.

Using data from a beverage manufacturer we show that our method produces more accurate forecasts than either the human or the machine on their own. Additionally, it reduces the human workload by ca. 50%, leaving the human forecasters more time for the cases in which they can add most value.

5 - Too Many Electric Vehicle Brands? Uncovering Market Demand Drivers in Alabama via Sparse Learning

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

We investigate key factors influencing battery electric vehicle (BEV) market demand using a high-dimensional panel dataset from Alabama, U.S. To identify the most significant predictors, we employ an enhanced sparse learning model based on Lasso regularization. Our findings reveal that a greater number of BEV brands in the market is significantly associated with reduced consumer demand. To assess the robustness and generalizability of the improved model, we further test it on two additional domains: GDP and the S&P 500 stock trading market.

6 - Enterprise Assets-Based Revenue Network (EARN) to Tokenize Oil and Gas Assets Through Fractional Ownership

Prasun Singh, Hitachi America, Ltd., R&D Division, Santa Clara, CA, United States, Ravigopal Vennelakanti, Takatoshi Ohara

Tokenizing real-world assets can revolutionize supply chains by linking physical industrial assets to digital ownership models. By converting these assets into non-fungible tokens (NFTs), they can be fractionally owned by a diverse pool of investors. This approach facilitates the creation of a dynamic digital marketplace, enabling greater access to capital and enhancing returns on infrastructure and equipment investments.

We propose an AI-enabled tokenization framework [Enterprise Assets-based Revenue Network (EARN)] to improve allocation visibility, responsiveness, liquidity, and valuation of oil and gas real-world assets (RWA). Fractional (NFT) owners get field operations incentives, such as tradable green energy power production for carbon credits. The framework provides operational AI and valuation insights. The EARN proof-of-concept can demonstrate the monetization of production oil and gas wells and incentivize disposal wells to decarbonize oilfield operations.

The objective function of the oilfield EARN framework is to maximize

$$U = \alpha V_t + \rho G_t + \beta L_t + \gamma S_t - \delta C_t - \epsilon R_t - \lambda n$$

Where:

U : Total Investor fractional possession utility of the oilfield RWA

V_t : Total value of tokenized assets

G_t : Total value added as operational incentives to revenue

L_t : Liquidity of the tokenized RWA (trading scale volume, investor participation, and trading efficiency – exchangeability or convertibility of token value)

S_t : Smart contract transparency

C_t : Costs for solution

R_t : Regulatory and legal risks

λ_n : Constraint or Penalty coefficients

The weights α, β, γ, δ, ε, ρ can be adjusted based on priorities such as maximizing liquidity value, reducing costs, and increase regulatory compliance.

WB07

Building A Level 3 A307

Transforming Operations Through Emerging Technologies

Contributed Session

Contributed

Chair: Manas Tripathi, XLRI- Xavier School of Management Delhi- NCR Campus India, Jhajjar, 124103, India

1 - Enhancing Grid Frequency Regulation with Shared Autonomous Electric Vehicles via a Two-Stage Robust Optimization Approach

XiaoFeng Li, Southeast University, China, Nanjing, China, People's Republic of, Xinlian Yu, shunfeng zhang

Shared autonomous electric vehicles (SAEVs) have the potential to provide both mobility services and grid stabilization services, such as frequency regulation, via Vehicle-to-Grid (V2G) technology. This dual functionality can enhance fleet utilization and profitability for operators while facilitating the integration of variable renewable energy sources. To address uncertainties in frequency regulation market prices and trip requests, we propose a data-driven two-stage robust optimization framework for the operation of SAEVs. In the first stage, the operator makes energy and capacity bids in the day-ahead regulation markets, considering uncertain regulation signals and travel demand. The goal is to maximize the platform's expected profit from selling primary frequency regulation to the grid and guarantees that market commitments are met at all times. In particular, the uncertainty market commitment is characterized via data driven deep reinforcement learning approach, which has the benefit of tractability. The second stage focuses on SAV operations, dynamically dispatching EVs to fulfill pre-committed regulation services and real-time trip requests.

Our computational study, based on real-world trip request data and actual frequency regulation prices, demonstrates that enabling SAVs to offer frequency regulation can significantly enhance the platform's expected profits while maintaining trip service quality. We also offer managerial insights into operational efficiency, particularly regarding fleet size, V2G station capacity, and trip request patterns. Our findings provide valuable insights for the design of V2G systems, potentially supporting the sustainable adoption of EVs and complementing renewable power growth by stabilizing the grid.

2 - Embodied AI Customer Service: Mitigating the Negative Impact of AI Identity Disclosure on Cognitive Trustworthiness and Emotional Affinity

Runge Zhu, Central University of Finance and Economics, Beijing, China, People's Republic of, Heyao Ma, Zhidi Shang

With the rapid advancement of artificial intelligence (AI), AI-based customer service is increasingly replacing human agents across various service domains. This research addresses three key questions: Can AI effectively resolve consumer inquiries? How do consumers react when informed the agent is AI rather than human? And, if negative reactions occur, what design strategies can mitigate them? In Study 1, we collected frequently asked questions from a major e-commerce platform and obtained responses from professional human agents and an AI tool. Participants evaluated these responses in a 2 (Response Source: Human vs. AI) × 2 (Label: Human vs. AI) between-subjects design. Results showed AI-generated responses were rated as more cognitively accurate and trustworthy and elicited emotional affinity comparable to human responses. However, when participants were informed the response was from AI, perceived trustworthiness and emotional affinity declined despite recognition of accuracy. In Study 2, we examined whether embodied design could attenuate this "AI label effect." A 2 (Label: Human vs. AI) × 2 (Embodiment: Present vs. Absent) experiment revealed that AI agents presented with visual avatars in immersive virtual reality significantly reduced the negative impact of AI disclosure on trustworthiness and emotional affinity compared to low embodiment (voice only). These findings provide critical insights for designing AI service agents in the evolving AI-driven landscape, guiding practitioners to develop strategies that enhance customer experience.

3 - Assessing the Potential of Drone-Based Stock Record Maintenance to Decrease Oversales in Retail

Toshi Protsch, ETH Zurich, Zurich, Switzerland, Oliver von Dzengelevski, Torbjørn Netland

Effective inventory management is crucial for the economic success of retail businesses. Inaccurate inventory records cause various challenges, one of which is unintentional oversales—selling a product beyond its available inventory. If the replenishment time is longer than the promised lead time, an oversale triggers a customer order that cannot be fulfilled. The ensuing delay or cancellation of orders may lead to increased operational costs, lost sales, and diminished customer satisfaction. This study examines the potential of drone-based systems for autonomous inventory record maintenance and their potential to mitigate oversales.

Based on sales and inventory data of a global retailer, we study the effect of drone implementation by comparing warehouses in which a drone-based autonomous stock counting system was implemented with comparable warehouses which continued to solely rely on manual stock counting. We employ a difference-in-differences methodology to study the effect of drone use on oversales, mediated by inventory record accuracy. Furthermore, we investigate the role of contextual moderators of this mediated relationship, including warehouse size, demand variability, item turnover rates, and customer access to inventory.

We contribute to the inventory management literature by providing empirical evidence on the effectiveness of stock record maintenance enabled by emerging technologies. We expect that drone-based stock record maintenance systems are particularly effective at preventing oversales in large warehouses, for fast-moving items,

with higher demand variability, and in sections inaccessible to customers. Our study provides practical guidance for retailers considering investments in autonomous record maintenance using emerging technologies.

4 - Digital Verification of Experience Certificates, Academic Credentials and Community Service from Source Organizations Using a Blockchain Based System

Sonal Daulatkar, Prin. L.N. Welinkar Institute of Management Development and Research, Mumbai, India, Tanmay Gurav, Jagdish Kharatmol, Nishil Visawadia, Sagar Deshmukh

Every candidate's background verification entails verification from the source organization across the world – corporate bodies for experience certificates, schools/universities for academic credentials and non-profit organizations for community service. This involves multiple stakeholders at different levels in the process including embassies of different countries for international candidates. Existing third-party businesses having liaison with the source organization charge heavily. This also introduces delays in the verification process and may also introduce human error while inserting marks into a database apart from the possibility of fraud. This project proposes a de-centralized block-chain based multi stakeholder platform for verification of documents as a solution. After a blockchain-based tokenized validation from the source organization, the verified document by a candidate is viewable on a web-based platform accessible to all concurrent and future stakeholders. The student community can apply globally to avail quality education while job applicants can apply for work visas with a lower verification cost.

5 - Generative AI in Action: Strategic Adoption and Performance Implications for Firms

Manas Tripathi, XLRI- Xavier School of Management Delhi- NCR Campus India, Jhajjar, India

Generative artificial intelligence (Gen-AI) has the potential to transform how firms create and capture value. Firms deploy Gen-AI technologies in content creation, product development, and customer service. Gen-AI enhances speed and creativity in research and development, promoting product and service innovation. Gen-AI signals innovation leadership to the customers and investors, shaping the competitive positioning of the firm. Prior research has emphasized the role of Gen-AI in enhancing innovation and operational efficiency, but empirical evidence to demonstrate its impact on firm performance remains limited. This study addresses this gap by examining the impact of Gen-AI announcements on firm performance using event study methodology. The event is the announcement of a Gen-AI initiative by a firm, and the event study measures the stock price performance of the firm in a defined event window around the event announcement date. The dataset consists of Gen-AI announcements post-2022 of US firms listed on the NYSE and NASDAQ. The study examines the short-term and long-term impact of Gen-AI announcements on the stock prices of US firms using the market model. Preliminary insights suggest that Gen-AI announcements have a statistically significant positive impact on firm stock prices, indicating investor optimism about the potential of Gen-AI to drive operational efficiency and innovative business models. The study contributes to the literature on artificial intelligence in strategic management by providing empirical evidence of the market impact of Gen-AI announcements. It offers early guidelines for investors, executives, and managers exploring the Gen-AI frontiers.

WB08

Building A Level 3 A310

Streaming and Digital Media Platforms

Invited Session

Information Systems

Chair: Maya Mudambi, 01040

Co-Chair: Yifei Wang, University of New Hampshire, Durham, NH, United States

1 - Navigating the Attention Economy: Content Strategy Trade-Offs for Live Streaming Entrepreneurs

Sung Hyun Kwon, University of Maryland, College Park, MD, United States, Il-Horn Hann

Live streaming entrepreneurs continuously decide whether to explore new content, exploit past successes, learn from competitors, or follow audience feedback. Understanding how content creators make these decisions and how audiences respond is critical as the streaming market intensifies. We examine these content strategy trade-offs among established streamers in a major live streaming platform by developing a framework that combines natural language processing with March's (1991) exploration-exploitation theory to quantify content strategies using streaming titles and game choices. We apply BLP structural demand estimation to analyze how these supplier decisions affect viewer utility and engagement. This study addresses how digital entrepreneurs navigate content strategy decisions when facing ephemeral content value and immediate audience feedback. By quantifying creative strategies in user-generated content markets, we provide insights into entrepreneurial decision-making in attention economies where traditional product development cycles do not apply. The framework offers a systematic approach to measuring content differentiation strategies, with implications for streaming entrepreneurs, platform design, and understanding supplier behavior in digital creative markets.

2 - Playlist-Induced Spillovers: Examining on- and off-Platform Outcomes for Emerging Artists on Spotify

Jose Eduardo Oros Chavarria, Carnegie Mellon University, Pittsburgh, PA, United States

Digitization and streaming platforms-based activities can create opportunities for artists at all stages of their careers. This paper studies the extended multiplier effects of editorial playlisting on both on-platform and off-platform outcomes for emerging, independent artists. Utilizing a unique dataset from Spotify's Playlist pitching tool, a feature that allows artists to pitch their tracks to music editors for playlist consideration, we employ a matched difference-in-differences methodology to compare outcomes between artists who were featured on playlists and those who were not. Our results show that emerging independent artists featured on Spotify's editorial playlists not only experience a significant increase in stream counts but also have a greater likelihood of going on tour. This research emphasizes the role of digital platforms as potent discovery mechanisms that enable fan engagement, in turn driving off-platform career trajectories including live event opportunities.

3 - The Nexus of Embodiment: Digital Agents in Live Streaming Contexts

Peiyan Yu, University of Georgia, Athens, GA, United States

Advances in generative technologies have enabled the rise of digital humans, embodied agents with lifelike appearance and expressive voice, who are increasingly used in advertising and sales. These agents offer a scalable and cost effective alternative to human influencers. Unlike traditional media figures, they can remain available around the clock and convey emotions, personality, and interactive behaviors that foster audience engagement. This study examines whether embodied agents can achieve performance comparable to humans in live streaming commerce and whether their effectiveness depends on how well they align with brand identity. We conduct an online experiment in a live streaming setting, varying agent features in terms of visual style and language use, and compare embodied agents to non embodied ones. We assess how the fit between brand and agent influences viewer engagement, brand perception, and purchase intent. The findings inform how brands can strategically deploy generative agents to enhance consumer connection and marketing effectiveness.

4 - The Impact of Realism and AI Disclosure on Virtual Influencer Effectiveness: a Large Field Experiment

Meixian Wang, Temple University, Philadelphia, PA, United States, Jaehwuen Jung, Ravi Bapna

Influencer marketing has become an essential component for brand strategies, and the adoption of AI-generated virtual influencers is accelerating with advancements in generative AI. However, these emerging practices raise challenges related to transparency and ethical concerns. By conducting a large field experiment involving over 1.8 million consumers, we examine the interplay between virtual influencers' anthropomorphism levels and the disclosure of their AI identity. The results show that virtual influencers with higher anthropomorphism levels enhance engagement metrics, while AI identity disclosure reduces link clicks and three-second video plays, particularly for highly realistic virtual influencers. Our underlying mechanism analysis based on an online experiment reveals that this reduction is driven by "expectation violation effect", where disclosure violates consumers' expectations and evokes negative feelings. Importantly, the negative effects reverse for consumers with prior experience interacting with virtual influencers—among this group, highly realistic virtual influencers with disclosure actually lead to greater engagement than their less anthropomorphic counterparts. Our findings provide theoretical and practical insights into the design and development of AI agents, emphasizing the need to strategically manage anthropomorphism and transparency to optimize consumer engagement.

WB09

Building A Level 3 A311

Data-Driven Retail Operations

Invited Session

MSOM: Service Operations

Chair: Nur Kaynar, Cornell University, Ithaca, NY, United States

Co-Chair: Ziwei Zhu, Cornell University, Ithaca, NY, United States

1 - Causal Product Networks: a Data-Driven Methodology for Modeling Basket-Shopping Consumer Behavior

Ziwei Zhu, Cornell University, Ithaca, NY, United States, Vishal Gaur, Nur Kaynar

The purchase of one product affects the purchases of others in a shopping basket. Discovering and measuring these effects across an exponentially large number of possible interactions among numerous products in different categories is extremely challenging. We propose a new causal discovery approach to learn the underlying causal structure among product purchases from observational shopping basket data, filter out non-causal correlations, and construct a causal product network (CPN) to describe these latent interactions. We validate this approach and demonstrate its value by utilizing a large-scale basket-shopping dataset. Our main results are as follows. (1) We show that CPNs represent interactions among products in a shopping basket much more accurately than other candidate network structures of hypothetical consumer behavior, such as a complete network or a correlations-based network. (2) Our unrestricted CPN model, which allows product-level relationships, more accurately fits basket-shopping data than restricted specifications of product interactions used in the previous literature, such as category-level interactions, without requiring significantly more parameters. (3) Comparing brick-and-mortar and online channels, we discover that the brick-and-mortar channel exhibits denser causal connections among product purchases in a basket. (4) Finally, using the constructed CPNs, we demonstrate their application in a multi-category assortment optimization problem and find that our model outperforms a benchmark multinomial logit model that treats each category independently by 20%-42% in total sales. Our results make a significant advancement towards using large-scale transaction data for modeling shopping behavior, generating new managerial insights, and optimizing decisions.

2 - The Role of Store Brands in Mitigating Operational Disruptions

Yasin Alan, Vanderbilt University, Nashville, TN, United States, Xabier Barriola, William Schmidt

We study the role of store brands on retailers' performance following operational disruptions. We exploit exogenous shocks in a difference-in-difference-in-differences regression specification that estimates the impact on grocery category sales for stores with a high (compared to a low) proportion of store brand market share. Our findings suggest that having strong store brands allows retailers to better manage operational disruptions, as retailers have greater control over the supply and pricing of store brand products.

WB10

Building A Level 3 A312

Pricing, Contracts, and Incentive Design in Platform Commerce

Contributed Session

Contributed

Chair: Xiang Ji, University of Science and Technology of China, Hefei, N/A, China, People's Republic of

1 - Optimal Subscription Box Design for Online Subscription Service

Renjun Hu, University of Michigan, Ann Arbor, MI, United States, Lennart Baardman, Hyun-Soo Ahn

Subscription services, with their fixed revenue streams per customer, critically depend on minimizing churn. This research introduces a framework for optimal subscription box design aimed at enhancing customer retention. We begin by establishing prior knowledge of customer preferences using machine learning models leveraging rich customer features. These initial beliefs are then dynamically updated through diverse forms of customer feedback, ranging from explicit product ratings to implicit consumption signals and retention decisions.

A key challenge is that learning true customer preferences is costly and ideally achieved within the first few interactions. Our model quantifies the value of different feedback types, recognizing that richer, more granular information significantly accelerates preference identification and allows for more precise box tailoring. This understanding of feedback value is foundational to designing effective mechanisms that incentivize customers to provide more detailed insights. By strategically motivating feedback, we can create a virtuous cycle: improved learning leads to better-personalized boxes, which in turn reduces churn and provides further opportunities for preference refinement.

This research has direct real-world applications and is being developed in collaboration with Hungryroot, a leading online grocery subscription business. The proposed models are validated using real customer data, and we are actively designing experiments to test mechanisms for incentivizing valuable feedback, demonstrating a practical pathway to improving customer satisfaction and lifetime value in subscription services.

2 - Returning Credits Versus Returning Cash in a Competitive Environment

Yiji CAI, Southern University of Science and Technology, Shenzhen, China, People's Republic of, Kanglin CHEN, Xiaoshuai Fan

Consumers often receive either credit or cash refunds for product returns or service cancellations. When receiving the credit, consumers enjoy price deductions in future purchases if they remember to redeem within the expiration term. Receiving cash offers immediate freedom in fund allocation as desired. These refund policies significantly influence consumer behavior and firm profitability. Thus, this study delves into the optimal design of refund policies, specifically examining the choice between refunding credit and refunding cash in a competitive market setting. Our investigation centers on the dynamic interactions between consumers and firms, considering two competing firms offering horizontally differentiated products over time. Employing a model of the repeated interactions across two periods, we analyze the impact of consumers' forgetfulness rates, cash freedom, and exogenous return rates on firms' pricing and refund decisions. Our analysis reveals several noteworthy insights. First, equilibrium choices regarding refunding cash or credit are symmetric, contingent upon consumer perceptions of cash freedom and forgetfulness rates. Notably, a prisoner's dilemma scenario arises when credit refunds dominate due to weak cash freedom and moderate forgetfulness among consumers. Second, cash refunds consistently offer full reimbursement, whereas credit refunds tend to provide partial reimbursements, particularly when forgetfulness rates are low. Finally, despite a decrease in consumer surplus with higher return rates, firms' profits and overall social welfare may increase under cash refund equilibrium.

3 - Selling a Product Line Through a Risk-Averse Platform

Xiang Ji, University of Science and Technology of China, Hefei, China, People's Republic of

This paper analyzes how a manufacturer's product line design interacts with a risk-averse platform's selection of selling mode. The results show that, if the manufacturer's product line extension fee is moderate, the risk-averse platform chooses the agency contract when the commission rate is low, but chooses the wholesale contract when the commission rate is high.

WB11

Building A Level 3 A313

Learning, Matching, and Decision Intelligence

Invited Session

Revenue Management

Chair: Yunzong Xu, University of Illinois, Champaign, IL, United States

1 - Alternating Regret for Online Convex Optimization

Mengxiao Zhang, University of Iowa, IOWA CITY, IA, United States, Soumita Hait, Ping Li, Haipeng Luo

Motivated by alternating learning dynamics in two-player games, a recent work by Cevher et al. (2024) shows that $o(\sqrt{T})$ alternating regret is possible for any T -round adversarial Online Linear Optimization (OLO) problem, and left as an open question whether the same is true for general Online Convex Optimization (OCO). We answer this question in the affirmative by showing that the continuous Hedge algorithm achieves $O(d^{2/3}T^{1/3})$ alternating regret for any adversarial d -dimensional OCO problems. We show that this implies an alternating learning dynamic that finds a Nash equilibrium for any convex-concave zero-sum games or a coarse correlated equilibrium for any convex two-player general-sum games at a rate of $O(d^{2/3}/T^{2/3})$. To further improve the time complexity and/or the dimension dependence, we propose another simple algorithm, Follow-the-Regularized-Leader with a regularizer whose convex conjugate is 3rd-order smooth, for OCO with smooth and self-concordant loss functions (such as linear or quadratic losses). We instantiate our algorithm with different regularizers and show that, for example, when the decision set is the ℓ_2 -ball, our algorithm achieves $O(T^{2/5})$ alternating regret with no dimension dependence (and a better $O(T^{1/3})$ bound for quadratic losses). We complement our results by showing some algorithm specific alternating regret lower bounds, including a somewhat surprising $\Omega(\sqrt{T})$ lower bound for a Regret Matching variant that is widely used in alternating learning dynamics.

2 - Optimal Spatial Matching on a Circle

Hao Liang, University of Illinois, Champaign, IL, United States, Zizhuo Wang, Yunzong Xu

In this study, we investigate spatial matching on a circular road. We derive new analytical results for the expected average matching distance, leading to interesting insights into more general geometric settings. Our findings also provide useful implications for the design of optimal delayed matching strategies for ride-hailing platforms.

3 - Balancing the Optimality-Interpretability Trade-off in Feature-Based Pricing

Jingren Liu, National University of Singapore, Singapore, Singapore, Hanzhang Qin, Mabel C. Chou

We consider a feature-based pricing problem under the predict-then-optimize (PTO) paradigm. In this setting, understanding the impact of features on demand is important for guiding pricing decisions. Although advanced machine learning (ML) models have shown strong predictive power in capturing complex feature effects, their black-box nature often limits interpretability, which is unacceptable in high-stakes decision-making domains. Therefore, it is desirable to harness predictive power while enhancing the interpretability of ML models. In this paper, we propose a Teacher-Student Monotonic Neural Network (TSMNN) to predict feature effects. In this architecture, a standard neural network is trained to balance prediction accuracy with similarity to a monotonic neural network (MNN) that lies in a subspace with lower Rademacher complexity. Such a favorable subspace is found through a theoretical but intuitive analysis of the generalization behavior of MNNs. Our proposed framework is completed by using double machine learning (DML) to estimate the price sensitivity. Theoretically, we show that the proposed framework yields a finite-sample regret bound. The empirical performance on a real-world loan data set further demonstrates both the superiority and practicality of our framework: it produces a monotonic pricing policy at a small cost in optimality.

4 - Data-Driven Diffusion Approximations of Graph Markov Chains

Prem Talwai, MIT, Cambridge, MA, United States, Jose Blanchet, David Simchi-Levi

Diffusion approximations are ubiquitous in the study of optimal control, sequential experimentation, and hypothesis testing, where they have long provided an elegant mechanism to distill Markovian dynamics to their initial moments. However, these approximations are purely asymptotic and largely restricted to the study of one-dimensional dynamics of highly stylized stochastic systems, which often invoke strong parametric assumptions on sequential increments (e.g. Gaussian, multinomial, exponential). However, in many OM settings, these parametric assumptions are too strong, and decision makers are faced with the critical learning problem of building estimators of graph transition kernels using only sample path data. On dense graphs, statistical estimation can be prohibitively expensive, scaling exponentially with the size of the state space. In this paper, we develop a nonparametric Bayesian approach for approximating Markovian dynamics on planar graphs by a special type of two-dimensional time-changed diffusion, known as Brownian motion with darning. Our approaches hinges on the association of the Green's function of the Brownian approximation with the period matrix of a carefully constructed hyperbolic surface. Harnessing the rich theory of Teichmüller dynamics and large genus Weil-Petersson geometry, we study the nonasymptotic posterior contraction of our Bayesian estimator around the Green's kernel of a "darned" projection, and the convergence of the latter to its continuous counterpart in the large-scale limit. We also explore the application of the diffusion approximation to optimal stopping, where we demonstrate that it enables a drastic reduction in the problem's statistical complexity, which now only scales linearly in the number of states.

WB12

Building A Level 3 A314

Stochastic Power System Planning: Modeling and Software Advancements

Invited Session

ENRE: Electricity

Chair: Rachael Alfant, Sandia National Laboratories, Albuquerque, NM, United States

1 - Forestlib: Enabling Performant Optimization of Stochastic Programs For Power Grid Planning

Rachael Alfant, Sandia National Laboratories, Albuquerque, NM, United States

Generation and transmission expansion planning is critical to the operation and maintenance of power grids; however, solutions to these problems can be computationally intractable. This talk introduces ForestLib, an open-source Python package with stochastic programming modeling capabilities, which incorporates solvers that optimize the extensive form and utilizes decomposition via progressive hedging. ForestLib is distinguished by its support for scenario bundling, multi-fidelity modeling, and a rich solution pool manager. Tailored to solve expansion planning problems, it utilizes parallelism and enhancements to the progressive hedging algorithm in order to achieve computational speed ups. We provide an expansion planning exemplar to demonstrate its use and the efficacy of our approach.

2 - Alternative Optima for Resilient Energy Modeling

Matthew Viens, Sandia National Laboratory, Madison, WI, United States

Effective planning for the future of the electricity grid requires improving our ability to treat complex capacity expansion planning problems. We show how generating alternative optimal solutions enhances our understanding of models generally. In capacity expansion planning problems, we use this to explore resilience and robustness characteristics. Our methods decompose cost efficiency modeling and more granular power systems simulation into separate pieces that can apply specialized tools to each component. We also leverage our previous work on alternative solution generation in Benders Decomposition to generate alternatives naturally over investment stage variables rather than the entire variable space.

3 - Using Snoglude, a Parallelized Python Package, to Perform a Priority Branch-and-Bound Global Optimization Algorithm on Stochastic Power Planning Problems

Georgia Stinchfield, Carnegie Mellon University, Pittsburgh, PA, United States, Matthew Viens, William Hart, Carl Laird

Two-stage stochastic programs exhibit a decomposable structure that many stochastic programming algorithms rely on to ensure computational tractability. In this work, we decompose by scenario, building one model for each set of random parameter realizations; each model and parameter set is referred to as a scenario. Aggregating the

optimal objective values after solving each scenario *independently* yields a valid lower bound to the original problem, as we no longer require first-stage variables (that appear in every scenario) to be the same across all scenarios (i.e., dropping constraints to yield a relaxation). We exploit this trivially parallelizable relaxation within the context of a priority branch-and-bound tree, where we branch only on the first-stage variables; by solving all scenarios to global optimality, we implicitly branch on the second-stage variables.

In the application of stochastic power planning problems, offering operators more than one “near-optimal” solution is often valuable to ensure transparency and interpretability of results. Developing a set of alternative solutions within a branch-and-bound tree is logical, as the space can be methodologically explored in a divide-and-conquer manner, naturally traversing the entire space of the optimization problem via tree search. Conventional branching techniques typically aim to close the optimality gap by solving a minimum number of nodes; in this work, we aim to traverse the tree while maximizing the variety of high-quality solutions available at termination. Our Python package, SNoGloDe: Structured Nonlinear Global Decomposition, allows us to employ our algorithm within a flexible software framework, easily adding branching customizations and candidate generators.

4 - Decomposition with Trust Regions for Scalable Stochastic Planning

Leonard Göke, ETH Zurich, Zürich, Switzerland, Giovanni Sansavini

Future power and energy systems are subject to various uncertainties; in particular, renewable sources, namely wind and solar, are characterized by an uncertain supply fluctuating from hour to hour and year to year. Cross- and auto-correlations of weather time series are too complex to describe analytically. Therefore, optimization models planning future systems must deploy a data-driven approach and consider large samples of climatic conditions to give reliable results. However, including a sufficient sample of climate conditions leads to very large linear capacity expansion problems that are computationally intractable for off-the-shelf solvers. Against this background, previous research proposed Benders decomposition for stochastic capacity expansion problems. The first stage of the resulting two-stage problem decides on complicating variables, most importantly capacities. The second stage decides on the operation and adds different scenarios, for instance, for various weather conditions. The original Benders algorithm converges slowly for large linear capacity expansion problems. However, previous research showed significant improvements from stabilization, viz. restricting the solution space for the current iteration to the neighborhood of the current best solution. We investigate synergies of stabilization using a quadratic trust region with other refinements of Benders decomposition. In particular, we leverage the trust region to transfer refinements for small combinatorial mixed integer problems to large but linear stochastic problems. A particular focus is on advanced stabilization, cut management, and cut selection techniques. Preliminary results on advanced stabilization suggest a performance improvement by a factor of five.

WB13

Building A Level 3 A315

Optimization and Machine Learning

Contributed Session

Contributed

Chair: Minseo Lee, 24060

1 - Uncovering Key Success Factors in Leading Semiconductor Enterprises: a Hybrid Approach Using Data Envelopment Analysis and Machine Learning

Chen-Hsiang Hong, National Taiwan University, Taipei, Taiwan, Chialin Chen, Ruey-Shan Guo

The semiconductor industry is a critical enabler of technological innovation and economic development, particularly in areas such as wafer foundry, advanced process engineering, and integrated circuit (IC) design. However, the sector is now confronted with unprecedented uncertainty due to the deceleration of Moore's Law, geopolitical tensions, and disruptive supply chain realignments. This research introduces a hybrid methodological framework that integrates Data Envelopment Analysis (DEA) with machine learning to assess firm-level operational efficiency and enhance strategic foresight. In the first phase, network and dynamic DEA models—including the Malmquist productivity index—are applied to evaluate technical and managerial efficiency, identify performance frontiers, and derive actionable insights. The second phase leverages supervised machine learning techniques to forecast market demand and track innovation trajectories. By fusing DEA's diagnostic strengths with the predictive capabilities of machine learning, this study aims to support evidence-based decision-making in areas such as R&D investment, technology adoption, and resource allocation. Key questions addressed include: (1) How can DEA reveal opportunities for improving innovation and operational performance? (2) How can machine learning strengthen strategic forecasting in volatile markets? (3) How can this integrated approach improve organizational agility in the evolving semiconductor ecosystem?

2 - Multi-Objective Trajectory Generation for Offline Model-Based Optimization

Chanho Kim, KAIST, Daejeon, Korea, Republic of, Jong-Seok Lee

In offline model-based optimization (MBO), guided trajectory generation (GTG) methods construct and denoise design trajectories toward high-performing regions. However, existing GTG approaches handle only a single objective, limiting their applicability to problems with multiple conflicting goals. To address this gap, we propose Multi-Objective Trajectory Generation (MOTG), a framework that extends GTG to multi-objective optimization. MOTG first extracts improvement trajectories from an offline dataset, ensuring that each step monotonically improves at least one objective. Each trajectory is annotated with a preference vector that captures the relative trade-offs of its final outcome. Using these labeled trajectories, we train a conditional generative model—either a diffusion model with classifier-free guidance or a continuous GFlowNet—to learn the distribution of multi-objective improvement trajectories. This generator enables us to sample novel trajectories conditioned on desired objective preferences. Finally, we generate a pool of candidate trajectories from the trained model and use a proxy function to select high-fidelity design solutions among them. Experiments on several multi-objective optimization benchmarks show that MOTG discovers diverse, high-quality solutions beyond the support of the original data. It consistently covers more of the Pareto front than existing methods, enabling more flexible exploration of trade-offs in offline settings.

3 - Stochastic Dual Cone Gradient Descent with Adaptive Moments: a Robust Approach to Conflict-Free Multi-Objective Optimization

Dohyun Bu, KAIST, Suwon, Korea, Republic of, Jong-Seok Lee

Multi-objective optimization in deep learning is often complicated by conflicting gradients among tasks and stochastic noise, which can hinder a model's ability to achieve well-balanced performance across objectives. Prior works have attempted to address these issues by either applying deterministic gradient projection techniques to resolve conflicts or introducing stochastic gradient corrections to reduce bias. However, these methods often struggle to simultaneously ensure conflict resolution and maintain robust convergence in the presence of noisy gradients. To overcome these challenges, we propose Stochastic Dual Cone Gradient Descent with Adaptive Moments, a novel optimization algorithm that combines conflict-free update directions with noise-robust gradient estimation. Specifically, our approach projects each stochastic gradient update onto the dual cone spanned by the task-specific gradients, ensuring that every update step is a descent direction for all objectives. In addition, we integrate an Adam-style adaptive moment estimator to track each task's gradient, yielding unbiased estimates of each task's true gradient from noisy samples. As a result, the proposed method guarantees conflict-free update directions in expectation and promotes stable convergence under stochastic conditions. We empirically validate the effectiveness of our approach by training Physics-Informed Neural Networks on representative partial differential equation benchmarks, where it achieves improved convergence stability and well-balanced performance across objectives.

4 - Two-Stage Decision Tree Optimization Using a GCN-Based Approach

Jose Navarro, University of South Florida, Tampa, FL, United States, Kaixun Hua, Alex Savachkin

We propose a Graph Convolutional Network (GCN) framework to approximate optimal decision tree (ODT) solutions as two-stage Mixed Integer Programming (MIP) problems. This approach utilizes advances in learning-to-optimize and knowledge distillation, featuring a dual GCN architecture: an encoder that learns decision tree structures from first-stage MIP variables and constraints, and a decoder for sample-specific classification. The encoder captures patterns of solver-optimized trees, while the decoder reconstructs leaf assignments, enabling efficient inference with interpretability. Preliminary results demonstrate that the GCN model accurately reconstructs solver outputs and replicates the classification behaviors of traditional MIP solvers.

5 - Formulating Transformers in Mixed-Integer Programming For Sequential Decision-Making

Minseo Lee, Virginia Tech, Blacksburg, VA, United States, Esra Toy

The transformer architecture has demonstrated outstanding performance across domains such as natural language processing, computer vision, and time-series analysis. This paper presents a novel mathematical programming formulation of the transformer, with a focus on integrating it into mixed-integer optimization (MIP) models for decision-making tasks. By explicitly modeling attention mechanisms and feedforward networks, we offer a formulation that captures the nonlinear dependencies inherent in sequential data while enabling provable optimality. We address key challenges, including the non-convexity of the softmax function, through piecewise linear approximations and decomposition strategies. Computational experiments validate substantial efficiency gains and improved solution quality over direct formulations. Furthermore, we demonstrate how this approach can be embedded in broader optimization problems, such as the resource-constrained shortest path problem, to leverage the transformer's representational power in complex combinatorial settings.

WB14

Building A Level 3 A316

Network Models in Optimization and Their Applications

Invited Session

Computing Society

Chair: Yajun Lu, Jacksonville State University, Jacksonville, AL, United States

1 - An Enhancement of Mixed-Integer Linear Programming Approaches for Resource Allocation in Project Networks

Norbert Trautmann, University of Bern, Bern, Switzerland, Nicklas Klein

An activity-on-node project network consists of nodes representing the project activities and edges representing the prescribed precedence relations between these activities. Executing an activity requires a prescribed amount of time and specific resources of different types. At any given time, the total consumption of each resource type cannot exceed its capacity. We consider the resource-constrained project scheduling problem (RCPSP), which consists in allocating the resources to the project activities over time in order to minimize the time-to-market, i.e. the project duration. A variety of formulations of the RCPSP as a mixed-integer linear programming model have been discussed in the literature. We propose an enhancement to existing models by incorporating a redundant constraint that evaluates the workload to be processed before and after the start of each activity, respectively. Our computational analysis on a standard test set from the literature confirms that this constraint considerably improves the relaxation-based lower bound.

2 - Induced Conflict Formulations for Second-Order K-Clubs

Yajun Lu, Jacksonville State University, Jacksonville, AL, United States, Baski Balasundaram

A k-club is a clique relaxation defined as a vertex subset inducing a subgraph with diameter at most k, originally introduced to model cohesive subgroups in social networks and tightly knit clusters in graph-based data mining. However, k-clubs can be fragile for $k \geq 2$, potentially creating subgraphs with low connectivity, such as those containing a cut vertex. This observation led various authors to introduce second-order k-clubs that require additional properties that endow the induced subgraph with fault tolerance and/or increased vertex connectivity. This talk introduces new induced conflict formulations for several second-order k-club models, offering the very first integer programming formulations for some of them. We compare the new formulations against existing formulations from the recent literature where available, investigate the effectiveness of the proposed formulation through theoretical and empirical means, and report encouraging computational evidence that this approach is favorable over existing approaches for challenging optimization problem instances.

WB15

Building A Level 4 A401

Location and Routing

Contributed Session

Contributed

Chair: Arun Akkinepally, Massachusetts Institute of Technology, Cambridge, MA, United States

1 - Modelling Gradual and Cooperative Coverage in Location Models

Mumtaz Karatas, Wright State University, Dayton, OH, United States, Levent Eriskin, Ertan Yakici

Location problems are a fundamental area of study in operations research and decision science, with wide-ranging applications spanning facility location, network design, emergency response, transportation, and defense. One critical aspect of location problems is the coverage of demand or service areas by facilities, which can have significant economic, social, and environmental implications. Conventional models often assume a binary or deterministic coverage, where a facility either fully covers a demand point or does not cover it at all. Such simplifications fail to capture the full complexity of real-world scenarios. This study explores the modeling challenges and solution approaches for location problems in the presence of gradual and cooperative coverage, where multiple facilities provide partial and joint coverage to demand points. We present mathematical formulations, linearization and approximation techniques and solution approaches for a range of models in retail operations and defense.

2 - A Semi-Desirable Modeling Approach for the Vertipoint Location Problem

Zahra Zare, Wright State University, Dayton, OH, United States, Mumtaz Karatas, Levent Eriskin

Abstract

In this study, we focus on the problem of locating vertipoints in the context of Urban Air Mobility (UAM). Vertipoints work as hubs to air-taxis and eVTOL (electric Vertical Take-Off and Landing aircraft) vehicles where passengers start by traveling to a vertipoint, fly between vertipoints using those vehicles, and then complete their trips using ground transportation. Vertipoints not only offer benefits like faster travel but also bring challenges like noise, visual pollution, and safety risks. To account for these conflicting effects, we model the problem as a semi-desirable hub location problem with both push and pull objectives. This approach enables planners incorporate both the positive impacts (better accessibility and connectivity) and negative impacts (disturbance to communities) when designing vertipoint networks. To tackle this problem, we first develop a weighted goal programming model to determine compromise solutions that balances total transportation costs and negative effects on communities. We model the disutility as the degree of impact vertipoints have on nearby neighborhoods using a distance-based function. We demonstrate the applicability of our formulation on a case study for Ohio, USA and perform a series of sensitivity analysis runs to explore the robustness of the solutions.

3 - Data-Driven Stochastic Vehicle Routing Problems with Deadlines

Shanshan Wang, Beihang University, Beijing (北京), China, People's Republic of, Erick Delage, Leandro Coelho

Vehicle routing problems (VRPs) with deadlines have received significant attention around the world. Motivated by a real-world food delivery problem, we assume that the travel time depends on the routing decisions, and study a data-driven stochastic VRP with deadlines and endogenous uncertainty. We use the non-parametric approaches, including k-nearest neighbor (kNN) and kernel density estimation (KDE), to estimate the decision-dependent probability distribution of travel time. To solve the resulting problem efficiently, we employ a logic-based Benders decomposition (LBBD) algorithm with several algorithmic enhancements. In particular, we propose a novel family of optimality cuts that includes the expected delay for all the subroutes. Moreover, we solve a total travel cost minimization problem to warm-start the algorithm. We also use a local search procedure to improve the current routing decision and propose a machine learning-based lower bound heuristic to efficiently solve

problems of realistic size. A practical case study for a food delivery routing problem using real-world data is conducted to show the efficiency of the proposed techniques and the advantage of the data-driven stochastic VRP in reducing the expected delay.

4 - Routing Autonomous Vehicles to Measure Pavement Friction in a Network

Caio Vitor Beojone, University of Illinois at Urbana-Champaign, Urbana, IL, United States, Yanfeng Ouyang

Pavement friction is critical for traffic safety. It is heavily associated with road crashes, notably when weather conditions are not ideal, making roads wet, snowy, or icy. However, accurate pavement friction is still challenging, costly, and slow in most critical spots. Hence, we propose a strategy to perform dynamic pavement friction measurements in a scenario where the authorities compile measurements from uncontrolled regular vehicles (RVs) and controlled autonomous vehicles (AVs). The strategy compiles measurements from different sources with varying methods by weighting their precision and grouping them into a single standardized distribution. The compiled measurements provide error estimates about the number of crashes associated with different arcs in the network. These estimates guide the routing of AVs, which try to minimize the errors of pavement friction measurements by covering the most critical spots in the network. A rolling horizon framework uses new incoming measurements to update routing decisions. This strategy was implemented in a prototype simulation based on Champaign County, reproducing realistic traffic counts for uncontrolled vehicles. Initial results showed that relying solely on uncontrolled RVs provides a sublinear convergence on pavement friction measurement precision. A couple of controlled AVs could cover about 20% of the arcs in the simulated network in three hours at a county level, and, compared with the exclusive use of RVs, they improved estimates by an additional 25%. Finally, further results show that, for planning horizons longer than 10 minutes, improvements in estimates depend mainly on the AV fleet size.

5 - Charge Planning for a Dedicated Electric Delivery Fleet

Arun Akkinipally, Amazon, Austin, TX, United States, Rami Sleiman, Andre Snoeck

We introduce Amazon's Charge Planner, an optimization algorithm designed to efficiently manage the charging of electric delivery vehicles, minimizing energy costs while ensuring operational readiness. As Amazon scales towards deploying 100k electric vehicles by 2030, effective charge planning becomes crucial for successful fleet electrification. We formulate charge planner as a MILP that optimizes for the charge scheduling while respecting infrastructure constraints, vehicle-specific charging requirements, and grid-level power constraints. The planner optimizes across multiple dimensions, including consumption-based spot costs and power-based demand charges, while ensuring that vehicles meet their target energy levels for next-day deliveries. We also enhance the approach to handle practical considerations like—changes in problem state and loss of charger connectivity under power-constraints. The results from simulations demonstrate that the charge planner achieves cost savings of 30-40% compared to baseline charging strategies. Finally, we also discuss future enhancements —like parking assignment and multi-day optimization.

WB17

Building A Level 4 A403

Building Industry Academic Collaborations for Innovations

Panel Session

Practice Track

Co-Chair: Saeid Motevali, Georgia State University, Atlanta, GA, United States

1 - Moderator Panelist

Saeid Motevali, Georgia State University, Atlanta, GA, United States

Strong partnerships between academia and industry can drive innovation, create actionable insights, and deliver meaningful impact. This panel highlights successful models and hard-earned lessons from professionals who have worked across sectors to bridge research and application. Panelists will share strategies for initiating collaboration, aligning incentives, and sustaining momentum across project lifecycles.

The conversation will focus on practical frameworks for building trust, negotiating intellectual property, and balancing academic rigor with business needs. Topics may include applied research programs, joint labs, embedded internships, and data-sharing partnerships. Panelists will speak from experience rather than company-specific examples, offering guidance that applies broadly across sectors and disciplines.

This interactive session invites questions from the audience and encourages thoughtful exchange around mutual benefit, innovation pipelines, and shared goals. Whether you're an academic seeking industry engagement or a practitioner navigating university collaboration, this panel offers actionable insights and new ways to connect ideas with impact.

2 - Panelist

Evan Charles, GLG, Atlanta, GA, United States

3 - Panelist

Hong Singhanian, Alimentation Couche-Tard, Marietta, GA, United States

WB18

Building A Level 4 A404

Logistics and Distribution Networks

Contributed Session

Contributed

Chair: Sina Aghakhani

1 - Optimization of Replenishment and Picking Processes in Robotic Mobile Fulfillment Systems with Multiple Picking Stations

Takashi Irohara, Sophia University, Tokyo, Japan, Shinnosuke KAWAHARA

In recent years, robotic mobile fulfillment systems (RMFS) have significantly enhanced the efficiency of the picking process. These systems employ robots to transport racks containing Stock Keeping Units (SKUs) to picking stations, where static pickers are utilized. The frequency of visits to workstations serves as a crucial metric for evaluating the system's effectiveness in terms of cost reduction and makespan compression. Previous studies have demonstrated that warehouses equipped with multiple picking stations can reduce the makespan of picking operations; however, these studies often overlook the tasks performed at replenishment stations. Replenishment activities have largely been neglected in this domain. Consequently, this study aims to minimize the total number of visits to multiple picking stations and a single replenishment station by optimizing both picking and replenishment processes. The proposed model is formulated as an integer linear programming problem. To validate the efficacy of the synchronized model, the outcomes of optimizing solely the picking processes and solely the replenishment processes are compared with those of the integrated model. The integrated model results in fewer visits to the replenishment station compared to the model optimizing only picking processes. Similarly, it results in significantly fewer visits to the picking stations compared to the model optimizing only replenishment processes. Numerical experiments demonstrate that simultaneous optimization of picking and replenishment processes can achieve a reduction in the total number of visits to all workstations, even when multiple picking stations are employed.

2 - Supply Chain Design Principles for SAF Production in Switzerland Using Anylogistix Network Optimization

Marco Aebi, Zurich University of applied Sciences, Winterthur, Switzerland, Maike Scherrer, Joel Weingart

Sustainable aviation fuels (SAF) are one option to reduce GHG emissions in the hard-to-abate aviation sector in the medium term. Among the different pathways that exist for SAF production, experts expect Power-to-Liquid (PtL) to be one of the most relevant ones in the future as it does not rely on limited resources. This pathway uses large amounts of (green) hydrogen and CO₂ as feedstock. As the supply of these feedstocks has not yet been scaled to the necessary quantity, many questions regarding the design of a SAF supply chain remain unanswered.

Using the simulation and optimization software AnyLogistix, this work aims to derive supply chain design principles for the SAF production from a transport perspective for the use-case of SAF production in Switzerland. Questions that will be addressed are for example if it is more beneficial from an environmental and economic perspective to produce SAF in close proximity to the feedstock or rather at the place of consumption, and what transport emissions occur in the SAF supply chain. The results enable informed decisions related to trade-offs in SAF production options. For this, a network optimization model including different modes of transport for the different compounds is used to optimize the SAF production network. Preliminary results show that it is cheaper to produce hydrogen directly at a power plant and then transport liquid hydrogen to the SAF plant compared to transmitting the electricity through the grid and producing hydrogen locally at the SAF plant.

3 - Maximizing Small Dairy Farm Revenues by Simultaneously Selecting Optimal Consolidation Depot Location and Routing Under Co-Mingling Constraints with Heterogeneous Vehicles and Time Windows

Ashesh Kumar Sinha, West Virginia University, Morgantown, WV, United States, John Saldanha, Molly Hughes, Michael Harman

We analyze a milk logistics network problem with heterogeneous vehicles, small farms, consolidation centers and processors. Milk logistics challenges include pickup and milk-grade based co-mingling restrictions and short shelf-life. The primary theoretical drivers for the novel supply chain management solution are transportation economies of density and network economies of scale. Typically, small farms are unable to fill even a small truck and therefore pay less than truckload (LTL) rates to haul their milk to the processor. Relatedly, as LTL rates increase with distance. We propose a novel MILP formulation that models a scenario of a cooperative network of farms and consolidation depots where milk is commingled and aggregated for distribution to processors involving several intermingled decisions: (i) find the optimal number and locations of consolidation depots. (ii) assign farms to each depot. (iii) determine the optimal route to run a truck to each farm to pick up milk and then deliver the milk to the depot. Here the tanker will commingle the milk picked up from each farm before delivery to the depot. (iv) determine transportation decisions for the milk at a much more economical truckload (TL) rate to the processors once the depot that received enough milk.

4 - Integrated Evacuation and Relief Supply Network Design for Foreseen Disasters

Vala Rahmati, Southern Methodist University, Dallas, TX, United States, Halit Uster

This study examines emergency preparedness for foreseen disasters such as hurricanes and floods. Given known evacuee counts for each evacuation zone, we determine the locations of distribution centers (DCs) and shelters, then route both relief supplies and evacuees through transfer nodes on an integrated time-phased network, minimizing cost while meeting arrival deadlines. The problem is formulated as a mixed-integer program and solved with an enhanced Benders decomposition that derives optimality cuts via master-problem modifications and surrogate constraints, further accelerated by solution-modification heuristics and speed-up cuts. Tests on randomly generated instances identify the most effective enhancements, and a GIS-based case study using South Texas data illustrates how key parameters influence network design and real-time operations.

5 - Optimization-Based Consolidation Engine for Last Mile Delivery

Hao Zhou, Amazon, Bellevue, WA, United States

We present an optimization-based approach for the last-mile delivery stop planning problem, where packages must be consolidated into efficient vehicle stops while considering both operational constraints and delivery quality. Current industry solutions often rely on rule-based heuristics, which can lead to suboptimal consolidation and require constant maintenance as delivery patterns evolve. We introduce an Optimization-based Consolidation Engine (OCE) that simultaneously determines stop grouping and optimizes delivery sequences by minimizing total service time across stops. The mathematical model balances the trade-off between parking time overhead and walking time while incorporating critical operational constraints. To address potential data quality issues and capture local delivery patterns, we enhance the base model with a novel hints mechanism that leverages historical delivery data through a cross-entropy loss function. The model is implemented using an efficient local search algorithm suitable for large-scale operations. In initial simulation analysis and pilot testing, we demonstrate improvements in both consolidation quality metrics and service time efficiency, with perfect consolidation rates increasing by 1.53% and under-consolidation decreasing by 2.29%. The approach provides a robust framework for incorporating dynamic features and adapting to real-time conditions. This research contributes to the growing literature on last-mile logistics optimization by presenting a practical solution to a complex operational challenge faced by modern delivery networks.

6 - Awesm: Informing Energy-Aware Last Mile Strategic Network Design

Harry Birnbaum, Amazon, Seattle, WA, United States, Andre Snoeck, Nic Wayand, Daniel Merchan

Amazon co-founded The Climate Pledge in 2019. As part of that Pledge, Amazon also announced to bring 100,000 electric delivery vehicles on the road by 2030. Introducing EVs requires changes across planning horizons, from strategic network design decisions, such as vehicle-to-delivery station allocations, to day-of execution. We propose an approach that quantifies the impact of servicing a specific area with EVs to inform such strategic network design decisions. This Area with EV Suitability Model (AwESM) flexibly defines geographical areas and fleet compositions, making it applicable across multiple strategic planning horizons and use cases.

7 - Improving Linehaul Efficiency with Load Consolidation Models

Jeren Konak, Georgia Institute of Technology, Atlanta, GA, United States, Amira Hijazi, Sikai Cheng, Alan Erera

Designing efficient routes and consolidation strategies for ground transportation between sorting hubs is a major challenge in large-scale logistics networks. Packages move through the system according to primary and alternate flow rules in trailers or other ground freight transportation modes as specified by the load plan. In collaboration with UPS, we analyzed potential partial loads that could feasibly be consolidated using historical load data, yielding insights for future load planning. These consolidations lead to lower overall costs and congestion while meeting service commitments and business practices.

Given the magnitude of the linehaul network, we have approached this problem by creating two robust optimization models – Consolidation by Single Destination (CBD) and Consolidation by Single Origin (CBO). The CBD model groups all partial loads that share a common destination into clusters based on the origin of the load. The model then searched for feasible consolidations within each origin cluster before the trailer reaches the shared destination. On the other hand, the CBO model identifies consolidation opportunities for all partial loads that are leaving from the same origin to different destinations. Both models aim to minimize costs (transportation, package handling/sorting) and the impact on driver schedules while maintaining important business-prioritized constraints, such as service commitments and sort capacities. Our findings offer promising directions for enhancing efficiency in large-scale logistics operations.

8 - A Multi-Objective Optimization Model for Two-Echelon Hybrid Delivery Planning Using Trucks, Robots, and Crowdsourcing

Sina Aghakhani, Oklahoma State University, Stillwater, OK, United States, Mohammad Fili, Guiping Hu

The rapid growth of e-commerce has heightened the demand for last-mile delivery systems that are not only efficient and cost-effective but also environmentally sustainable. This study proposes a multi-objective mathematical model for a hybrid delivery planning framework that integrates trucks, autonomous delivery robots, and crowdsourced drivers to optimize logistics operations. The proposed two-tier model assigns trucks to transport packages to designated parking hubs, from which autonomous robots and crowdsourced drivers complete the final leg of delivery. The model aims to minimize the total delivery costs while also reducing overall operation time. We employ Non-dominated Sorting Genetic Algorithm II (NSGA-II) to find the pareto optimal solutions and compare the performance against the augmented ϵ -constraint method for small instances. Additionally, we introduce and evaluate the algorithm on a large-scale Brazilian E-commerce dataset in São Paulo.

WB19

Building A Level 4 A405

Experimental Design and Networks

Invited Session

Applied Probability Society

Chair: Shuangning Li, University of Chicago, Chicago, United States

1 - Advances in Budget-Split Experimentation for the Online Ads Marketplace

Shan Ba, LinkedIn, Mountain View, CA, United States

Classical A/B testing on online advertising platforms is often subject to bias, particularly due to cannibalization effects. Budget-Split Testing (BST) is widely regarded as the gold standard for mitigating these biases, as it controls the allocation of campaign budgets across different treatment variants. In this talk, we present recent methodological and system-level advancements that improve the design and analysis of BST. We examine how BST interacts with marketplace dynamics, explore the operational trade-offs of increasing the number of partitions (i.e., splits in BST), and assess the associated cost implications. Additionally, we identify various sources of bias in standard BST setups and introduce a Robust BST design that systematically eliminates partition-related biases. These enhancements make BST more scalable and reliable for high-stakes decision-making in optimizing the online ads marketplace.

2 - Enhancing Geo-Targeted Experimentation: Robust Methodologies for Measuring Advertising Effectiveness in Privacy-Constrained Environments

Malika Korganbekova, The University of Chicago Booth School of Business, Chicago, IL, United States, Jerry Chen, Faith Shin

This paper addresses the challenges of conducting advertising effectiveness experiments in privacy-restricted environments, where data limitations and stringent regulations hinder precise measurement. Platforms increasingly rely on geo-targeted experiments, such as cluster-randomized tests based on geolocations, to navigate these constraints. However, these tests face significant obstacles, including low statistical power, imprecise geolocation matching, and complications arising from dense urban clusters. A key challenge is **interference**, which occurs when individuals in treatment and control groups influence each other's outcomes, often due to overlapping geographies or shared advertising exposure. Interference distorts estimates of treatment effects and reduces the reliability of results.

We partner with a large online retailer in the US and propose a comprehensive framework to enhance the power and precision of geo-targeted experiments. Our methodology includes optimizing geo-split designs to increase sample size, applying synthetic control methods to create robust comparison groups, and introducing switchback testing to improve accuracy in dynamic settings. Clustered randomization is employed to group geographically close areas, minimizing interference while preserving granularity. Re-clustering techniques further balance treatment assignments by accounting for proximity and population density. Simulations are used to model the impact of interference and refine confidence intervals, while observational data aids in reducing measurement biases.

Through simulations, historical data analysis, and new experimental designs, we evaluate the effectiveness of these methods. Our findings contribute to the development of privacy-compliant experimentation frameworks that improve targeting accuracy, boost customer lifetime value, and enhance profitability. Beyond marketing, these approaches have broader applications, such as optimizing logistics and supply chain.

3 - Local Geometric Structure Detection in Random Graphs: Information-Theoretic Limits

Shuangping Li, Yale University, New Haven, CT, United States

We study the problem of detecting a planted random geometric subgraph within an Erdős-Rényi graph. We characterize the information-theoretic limits of detection as a function of the ambient dimension, identifying thresholds where geometric structure becomes distinguishable from randomness.

4 - Choosing the Better Bandit Algorithm Under Data Sharing: when Do A/B Experiments Work?

Jingyan Wang, Toyota Technological Institute at Chicago, Chicago, IL, United States, Shuangning Li, Chonghuan Wang

We study A/B experiments that are designed to compare the performance of two recommendation algorithms. Prior work has shown that the standard difference-in-means estimator is biased in estimating the global treatment effect (GTE) due to a particular form of interference between experimental units. Specifically, units under the treatment and control algorithms contribute to a shared pool of data that subsequently train both algorithms, resulting in interference between the two groups. The bias arising from this type of data sharing is known as "symbiosis bias". In this paper, we highlight that, for decision-making purposes, the sign of the GTE often matters more than its precise magnitude when selecting the better algorithm. We formalize this insight under a multi-armed bandit framework and theoretically characterize when the sign of the expected GTE estimate under data sharing aligns with or contradicts the sign of the true GTE. Our analysis identifies the level of exploration versus exploitation as a key determinant of how symbiosis bias impacts algorithm selection.

5 - Hypothesis Testing for Detecting Network Interference in Randomized Experiments

Christopher Harshaw, Columbia University, New York City, NY, United States, Yitan Wang, Chao Gao, Fredrik Sävje

From political science and economics to public health and corporate strategy, the randomized experiment is a widely used methodological tool for estimating causal effects. In order to meaningfully estimate causal effects, experimenters have to assume structure on the potential outcome functions. The simplest assumption is that of no-interference, although a growing body of work has provided various network interference models under which causal effects can be estimated. Still, this raises the question: to what extent can we test these assumptions? For example, can we detect interference when it is present in the experiment? Can we test between stronger and weaker models of network interference? To this end, a line of previous work has focused on controlling Type I error, but Type II error of these proposed methods remains less well understood.

In this work, we further investigate hypothesis testing frameworks for detecting interference. Our main contribution is a series of negative results. Roughly speaking, we show that there is no test for detecting interference which has small Type I and Type II error, even as the sample size increases. This finding has serious consequences for the epistemological nature of the no-interference and various network interference assumptions: namely, that they are -- in a rigorous sense -- unstable. On the positive side, we show that there exists a uniformly consistent test of no interference when the alternative hypotheses are restricted to a linear-in-means type model and the graph is sufficiently sparse.

WB20

Building A Level 4 A406

Multiserver-job Scheduling and Collusive Learning

Invited Session

Applied Probability Society

Chair: Izzy Grosz, Northwestern University, Evanston, IL, 60201, United States

1 - Half-Batch Multiserver-Job Scheduling

Ziyuan Wang, Northwestern University, Evanston, IL, United States

In large-scale computing systems such as high-performance clusters, jobs often demand heterogeneous server allocations: large jobs that occupy a substantial fraction of the servers are latency-sensitive, while small jobs fill in the remaining capacity to maintain utilisation. To model this dynamic, we introduce the half-batch multiserver-job (MSJ) model, a queueing model in which large jobs arrive according to a Poisson process and require all servers simultaneously, while small jobs, each needing only one server, are always available. We study the resulting trade-off between large-job response time and small-job throughput, define a two-dimensional performance criterion, and aim to characterize the Pareto-optimal frontier.

We analyse a practical family of scheduling policies, the $\$(k,n)\$$ -batch policies and their convex combinations, and propose that they form the Pareto frontier in the half-batch MSJ system. We establish partial structural results about the sample paths of optimal policies and outline the final result needed to complete the proof that convex combinations of $\$(k,n)\$$ -batch policies characterise the full Pareto-optimal frontier.

2 - Multiserver-Job Scheduling with Continuous Requirement Distribution

Heyuan Yao, Northwestern University, Evanston, IL, United States, Izzy Groszof, Jay Kowalik

Multiserver-job systems are a useful model for large-scale computing systems, where different jobs may require different amounts of resources. Existing literature assumes a discrete set of possible job requirements. To better model real-world systems, we study a model where the job requirement distribution is continuous.

In this talk, we will introduce the K-discretized Max-Weight policy, where we equally partition the capacity into K buckets and in each bucket round up the requirement. We prove a full-utilization result when the requirement distribution has a decreasing density. Moreover, we prove throughput-optimality when the requirement density is Lipschitz. We also prove equivalent results for the nonpreemptive K-discretized Markovian Service Rate policy. We also generalize to the adaptive-discretization setting and the multiresource-job setting.

3 - Collusive Pricing Under Oblivious Learning

Yuhang Wu, Columbia University, New York, NY, United States, Assaf Zeevi

There is a growing concern that independent pricing algorithms under competitive market conditions may learn to sustain near-collusive outcomes, a phenomenon called algorithmic collusion. Aside from the regulatory implications, this phenomenon also raises nontrivial questions on how firms should learn and make pricing decisions when it is possible to collude tacitly. To further our understanding of these problems, we propose a novel online multi-agent framework and study the market outcome when sellers are "oblivious" to competitor information. We first provide insights on how sellers' price exploration behaviors affect the market outcome and pinpoint sellers' optimal strategy from a revenue management perspective. Building on these insights, our analysis indicates that, contrary to some existing literature, collusive behaviors do not sustain very naturally. In fact, though significant excursions towards the collusive outcome are possible over finite horizons, the system always converges to the competitive outcome in the long run. We also compare oblivious sellers with "non-oblivious" sellers and illustrate the algorithmic dynamics between these different modeling choices. Supported by both theoretical and numerical evidence, we believe our results shed light on whether sellers should incorporate competitor information in their learning and pricing models. More broadly, our work contributes to the study of algorithmic collusion by showing how revenue management principles can both inform and structure its analysis, paving the way for a deeper understanding of its implications.

4 - Gradient-Based Algorithms for Machine Unlearning

Siqiao Mu, Northwestern University, Evanston, IL, United States, Diego Klabjan

Machine unlearning algorithms aim to efficiently remove data from a model without retraining it from scratch, in order to remove corrupted or outdated data or respect a user's "right to be forgotten." Certified machine unlearning is a strong theoretical guarantee based on differential privacy that quantifies the probabilistic indistinguishability between the unlearned model and the model retained on the retained data samples. We introduce new gradient-based algorithms for certified machine unlearning, emphasizing the importance of *black-box* algorithms that can be applied to trained models without costly precomputation.

5 - Multiserver-Job Response TIME Under Multilevel Scaling

Izzy Groszof, Northwestern University, Evanston, IL, United States

We study the multiserver-job setting in the load-focused multilevel scaling limit, where system load approaches capacity much faster than the growth of the number of servers n . We specifically consider the " 1 and n " system, where each job requires either one server or all n

servers. Within the multilevel scaling limit, we examine three regimes: load dominated by n -server jobs, 1 -server jobs, or balanced.

In each regime, we characterize the asymptotic growth rate of the boundary of the stability region and the scaled mean queue length. We numerically verify our asymptotic results against exact formulae.

WB21

Building A Level 4 A407

Learning and Optimization for Urban Logistics

Invited Session

TSL: Urban Transportation Planning and Modeling

Chair: Xinwu Qian, Rice University, Houston, TX, United States

1 - The Dynamic Stochastic TIME-Dependent Vehicle Routing Problem with Non-Recurring Congestion Events: a Reinforcement Learning Approach

Fernando Pilleux, Universidad Adolfo Ibañez, Santiago, Chile, Felipe Lagos, Stefano Novellani

In recent times, technological advances and the COVID-19 pandemic have driven a shift in consumer preferences, with a growing demand for faster delivery services. This transformation has led to a significant increase in the global e-commerce market and has forced companies to emphasize last-mile delivery. This study formulates a Dynamic Stochastic Time-Dependent Vehicle Routing Problem with Non-Recurrent Congestion (STDVRPTW-NRC), aiming to minimize the total operational cost. The problem arises from the need to accurately capture real-world traffic conditions using real data from the city of Chengdu, China. The model accounts for speeds and travel times that vary stochastically and depend on the time of day. Additionally, it introduces an extra layer of uncertainty through the inclusion of non-recurrent congestion events. The problem is formulated as a Markov Decision Process and solved using a reinforcement learning approach, specifically through an on-policy Monte Carlo control (MC) method with function approximation. This method employs a linear approximation of the value function, enabling it to effectively adapt to the uncertainty that unfolds during the execution of the problem. The proposed approach is compared with a state-of-the-art metaheuristic that provides a static solution to the problem. To evaluate the performance of these models, four different non-recurrent congestion scenarios are tested, using instances with an average of 150 and 250 customers. Across all scenarios evaluated, the on-policy MC method consistently outperformed benchmark policies, achieving improvements ranging from 11.03% to 41.26% in total costs.

2 - Multi-Agent Reinforcement Learning for Optimizing Relocation in One-Way Electric Car Sharing Systems

Wentao Zhao, University of Southern California, Los Angeles, CA, United States, Maged Dessouky

We study the dynamic relocation problem in an electric vehicle (EV) sharing system that provides on-demand mobility services. The objective of this research is to develop a high-quality, real-time relocation strategy for a widely distributed, large-scale shared EV network, capable of addressing stochastic customer demand and mitigating the computational challenges arising from the curse of dimensionality. To this end, we propose a decentralized two-stage decision-making scheme. In the first stage, we develop a multi-agent reinforcement learning framework that models each EV-sharing station as an independent agent responsible for determining its relocation requests. In the second stage, we transform the decentralized relocation requests into a feasible system-wide relocation strategy that satisfies operational constraints, such as fleet conservation, by solving an optimization problem. Finally, we evaluate the performance of the proposed method by comparing it against several optimization-based benchmarks in terms of profit and service quality, using real-world data from a major EV-sharing system.

3 - Infinite Drive In the City: Scalable Optimization of Dwpt Locations for Urban EV Charging Using Gmns

Yudai Honma, The University of Tokyo, Tokyo, Japan, Xuesong Zhou, Katsuhiro Hata, Takashi Oguchi

We develop a scalable optimization framework for deploying Dynamic Wireless Power Transfer (DWPT) infrastructure to support continuous electric vehicle (EV) operation in urban environments — a concept we call "Infinite Drive." Our mixed-integer programming model minimizes DWPT installation length while ensuring sufficient charging across all travel routes. It incorporates detailed traffic dynamics, including acceleration, deceleration, and signal-induced stopping behavior, and evaluates route-level energy balance using fine-grained segment analysis. Applied first to Kawagoe, Japan, the model identifies just 2,359 meters of DWPT — less than 2% of total road length — as sufficient to support full EV mobility. We extend the model to New York City using the GMNS dataset and find similarly effective results despite increased network complexity. The study demonstrates the feasibility and efficiency of DWPT infrastructure in diverse urban settings, offering a replicable and

data-driven tool for cities seeking to build plug-free, low-emission mobility systems. This research provides a foundation for scalable planning of dynamic EV charging and supports the broader transition to sustainable urban transportation.

4 - Enhancing Online Food Delivery with Transfer Points: a Decompose-Then-Optimize Approach via Hierarchical Reinforcement Learning

Xinwu Qian, Rice University, Houston, TX, United States, Xinyuan Zhang, Qi Luo

Online food delivery services can reduce operational costs and optimize efficiency by consolidating orders with similar origins, destinations, and time windows at intermediate transfer locations.

This research investigates the complexity inherent in the online food delivery problem with transfer (OFDP-T) and assesses how optimized routes and courier assignments involving transfer locations can enhance system delivery performances.

We propose a novel learning-based decompose-then-optimize framework to manage the exponentially growing problem size introduced by transfer and synchronization decisions and enable adaptive decision-making under uncertainty.

This proposed decomposition framework is enabled by a seamless integration between a first-step hierarchical reinforcement learning (HRL) model and the resulting second-step model that can be solved as a linear assignment problem (LAP).

Through comprehensive experiments based on real-world food delivery data, the study demonstrates that the combination of task-agnostic reward design and LAP-guided policy search significantly improve the baseline methods.

Our case study shows that task-agnostic reward design and LAP-guided policy search improve baseline performance by 27.2%, with the reward shaping alone boosting HRL by 37.5% and LAP-guided search adding 6.9%. Notably, even limited use of transfers can yield over 46.6% improvement in route efficiency and a 23.2% gain for remaining orders. This framework offers a deployable, real-time solution and actionable strategies for coordinating complex delivery operations and improving fleet utilization that will empower more sustainable and scalable food delivery systems

5 - Integrated Order Dispatching and Routing via Hierarchical Deep Reinforcement Learning for Last-Mile Delivery

Yida Xu, Tianjin University, Tianjin, China, People's Republic of, Henry Miao, Zhaofang Mao

In recent years, the increasing complexity of last-mile delivery has placed significant demands on the speed and accuracy of decision-making on delivery platforms. Traditional approaches that address the order dispatching and courier routing problems separately often result in sub-optimal solutions and lack coordination between these critical components. To address this gap, we formulate a double-level optimization problem integrating both order dispatching and courier routing, leveraging a dataset of last-mile pickup deliveries provided by Cainiao Logistics (LaDe, 2024). To solve this challenging problem, we propose a hierarchical deep reinforcement learning framework. For the upper-level order dispatching problem, we encode nodes and edges in a heterogeneous graph to distinguish courier nodes from order nodes. To capture the structural similarity of orders from different communities and the dissimilarity among neighboring orders, we design a Spatial Heterophily graph structure to optimize the efficiency of order dispatching. For the lower-level routing problem, a Dynamic Residual Graph Attention Network Encoder is proposed to capture both temporal and spatial features, along with a Look-Ahead Courier Personalized Decoder. Experiments with real-world data confirm the effectiveness of our proposed method. The results show a significant advantage over popular deep reinforcement learning methods and heuristic methods, indicating that our method can effectively help package delivery companies solve the real-time package pickup problem. This research provides a novel integrated approach to optimizing last-mile delivery operations, offering benefits for academics through innovative methodologies, for practitioners through enhanced operational efficiency, and for policymakers through insights into sustainable urban logistics solutions.

WB22

Building A Level 4 A408

Reinforcement Learning with Preference Feedback

Invited Session

Applied Probability Society

Chair: Qiaomin Xie, University of Wisconsin-Madison, Madison, United States

Co-Chair: Yixuan Zhang, University of Wisconsin-Madison, Madison, WI, United States

1 - Zeroth-Order Policy Gradient for Reinforcement Learning from Human Feedback Without Reward Inference

Qining Zhang, University of Michigan, Ann Arbor, MI, United States, Lei Ying

Reward inference (learning a reward model from human preferences) is a critical intermediate step in the Reinforcement Learning from Human Feedback (RLHF) pipeline for fine-tuning Large Language Models (LLMs). In practice, RLHF faces fundamental challenges such as distribution shift, reward model overfitting, and problem misspecification. An alternative approach is direct policy optimization without reward inference, such as Direct Preference Optimization (DPO), which provides a much simpler pipeline and has shown empirical success in LLM applications. However, DPO utilizes the closed-form expression between the optimal policy and the reward function, which is only suitable under the bandit setting or deterministic MDPs. This paper develops two RLHF algorithms without reward inference for general RL problems beyond bandits and deterministic MDPs, and general preference models beyond the Bradley-Terry model. The key idea is to estimate the local value function difference from human preferences and then approximate the policy gradient with a zeroth-order gradient approximator. For both algorithms, we establish polynomial convergence rates in terms of the number of policy gradient iterations, the number of trajectory samples, and human preference queries per iteration. Numerical experiments in stochastic environments validate the performance of our proposed algorithms, outperforming popular RLHF baselines such as DPO and PPO. Our paper shows there exist provably efficient methods to solve general RLHF problems without reward inference.

2 - Beyond Bradley-Terry Models: a General Preference Model for Language Model Alignment

Quanquan Gu, UCLA, Los Angeles, CA, United States

Modeling human preferences is crucial for aligning foundation models with human values. Traditional reward modeling methods, such as the Bradley-Terry (BT) reward model, fall short in expressiveness, particularly in addressing intransitive preferences. In this paper, we introduce preference embedding, an approach that embeds responses into a latent space to capture intricate preference structures efficiently, achieving linear query complexity. Additionally, we propose preference score-based General Preference Optimization (GPO), which generalizes reward-based reinforcement learning from human feedback (RLHF). Experimental results show that our General Preference embedding Model (GPM) consistently outperforms the BT reward model on the RewardBench benchmark and effectively models cyclic preferences where any BT reward model behaves like a random guess. Furthermore, evaluations on downstream tasks such as AlpacaEval2.0, following the language model post-training with GPO and our general preference model, reveal performance improvements over BT models. These findings indicate that our method may enhance the alignment of foundation models with nuanced human values.

3 - Faster Wind: Accelerating Iterative Best-of-N Distillation for LLM Alignment

Tong Yang, Carnegie Mellon University, Pittsburgh, PA, United States

Recent advances in aligning large language models with human preferences have corroborated the growing importance of best-of-N distillation (BOND). However, the iterative BOND algorithm is prohibitively expensive in practice due to the sample and computation inefficiency. This paper addresses the problem by revealing a unified game-theoretic connection between iterative BOND and self-play alignment, which unifies seemingly disparate algorithmic paradigms. Based on the connection, we establish a novel framework, WIN rate Dominance (WIND), with a series of efficient algorithms for regularized win rate dominance optimization that approximates iterative BOND in the parameter space. We provides provable sample efficiency guarantee for one of the WIND variant with the square loss objective. The experimental results confirm that our algorithm not only accelerates the computation, but also achieves superior sample efficiency compared to existing methods.

4 - Online Rlhf with Exploration

Jincheng Mei, Google, St Albans, United Kingdom, Bo Dai

Reinforcement learning (RL) has demonstrated great promise in aligning large language models (LLMs) with human preference, but how to collect samples efficiently for training is still the major bottlenecks. Concretely, while the principles of optimism or pessimism under uncertainty are well-established in standard RL to obtain more effective samples, a practically-implementable and theoretically-grounded form amenable to large language models is not yet available, as standard techniques for constructing confidence intervals become intractable under arbitrary policy parameterizations. In this talk, we introduce value-incentivized preference optimization (VPO), which regularizes the maximum-likelihood estimate of the reward function with the corresponding value function, as a theoretical guarantee implementation of optimism prior for exploration, matching the rates of their standard RL counterparts. Moreover, experiments on text summarization and dialog verify the practicality and effectiveness of VPO.

WB24

Building A Level 4 A410

Recent Advances in Min-Max Optimization

Invited Session

OPT: Nonlinear Optimization

Chair: Mert Gurbuzbalaban, Rutgers, New Brunswick, NJ, United States

Co-Chair: Necdet Aybat, Penn State University, University Park, PA, 16802, United States

1 - High-Probability Complexity Bounds for Stochastic Non-Convex Minimax Optimization

Mert Gurbuzbalaban, Rutgers University, Piscataway, NJ, United States, Yassine Laguel, Yasa Syed, Necdet Aybat

Stochastic smooth nonconvex minimax problems are prevalent in machine learning, e.g., GAN training, fair classification, and distributionally robust learning. Stochastic gradient descent ascent (GDA)-type methods are popular in practice due to their simplicity and single-loop nature. However, there is a significant gap between the theory and practice regarding high-probability complexity guarantees for these methods on stochastic nonconvex minimax problems. Existing high-probability bounds for GDA-type single-loop methods only apply to convex/concave minimax problems and to particular non-monotone variational inequality problems under some restrictive assumptions. In this work, we address this gap by providing state-of-the-art high-probability complexity guarantees for nonconvex/PL minimax problems corresponding to a smooth function that satisfies the PL-condition in the dual variable. We also present numerical results on a nonconvex/PL problem with synthetic data and on distributionally robust optimization problems with real data, illustrating our theoretical findings.

2 - A Stochastic Gradient Ascent Descent Method with Backtracking

Necdet Aybat, Penn State University, University Park, PA, United States, Mert Gurbuzbalaban, Qiushui Xu, Xuan Zhang

We propose a stochastic GDA method with backtracking (SGDA-B) to solve nonconvex-(strongly) concave minimax problems. SGDA-B is agnostic to the Lipschitz constant L , concavity modulus $\mu \geq 0$, and the variance bound of the unbiased stochastic gradient estimator. Within $\mathcal{O}(\epsilon^{-4} \log(1/p))$ and $\mathcal{O}(\epsilon^{-7} \log(1/p))$ stochastic gradient calls, SGDA-B can compute an ϵ -stationary point in terms of gradient-map norm and gradient norm of the Moreau envelope with probability $1-p$ when $\mu > 0$ and $\mu = 0$, respectively.

3 - AGDA+: Proximal Alternating Gradient Descent Ascent Method with A Nonmonotone Adaptive Step-Size Search for Nonconvex Minimax Problems

Qiushui Xu, Penn State University, State College, PA, United States, Xuan Zhang, Necdet Aybat

We consider double-regularized nonconvex-strongly concave (NCSC) minimax problems of the form $(P): \min_{\mathbf{x}} \max_{\mathbf{y}} g(\mathbf{x}) + f(\mathbf{y}) - h(\mathbf{y})$, where g, h are closed convex, f is L -smooth in $(\mathbf{x}, \mathbf{y})$ and strongly concave in $\mathbf{y}$. We propose a proximal alternating gradient descent ascent method `AGDA+` that can adaptively choose nonmonotone primal-dual stepsizes to compute an approximate stationary point for (P) without requiring the knowledge of the global Lipschitz constant L and the concavity modulus μ . Using a nonmonotone step-size search (backtracking) scheme, `AGDA+` stands out by its ability to exploit the local Lipschitz structure and eliminates the need for precise tuning of hyper-parameters. `AGDA+` achieves the optimal iteration complexity of $\mathcal{O}(\epsilon^{-2})$ and it is the first step-size search method for NCSC minimax problems that require only $\mathcal{O}(1)$ calls to ∇f per backtracking iteration. The numerical experiments demonstrate its robustness and efficiency.

4 - Risk-Sensitive Accelerated Gradient Methods with Biased Oracles

Yasa Syed, Rutgers University - New Brunswick, New Brunswick, NJ, United States, Mert Gurbuzbalaban, Necdet Aybat

We study generalized momentum methods (GMMs) for minimizing strongly convex, smooth functions in the presence of potentially biased stochastic gradient noise. This broad class of optimization algorithms includes gradient descent, the heavy ball method, and Nesterov's accelerated gradient method as special cases. Using the risk sensitivity (RS) index from robust control, we examine how step size and momentum parameters influence both convergence behavior and robustness to noise and bias. For quadratic objectives with independent and identically distributed Gaussian noise, we derive a closed-form expression for the RS index and establish large deviation results for the average suboptimality, along with high-probability bounds. We then extend these results to general strongly convex functions under a light-tailed noise assumption, deriving analogous RS index bounds, large deviation bounds, high-probability guarantees, and suboptimality estimates. Finally, we discuss the design of momentum methods that achieve optimal robustness properties in the presence of biased stochastic gradients.

WB25

Building A Level 4 A411

Transportation Systems

Contributed Session

Contributed

Chair: HUA LI, Yonsei university, Seoul, Korea, Republic of

1 - Traffic Reduction Potential of Autonomous Driving Freight Vehicles in Urban Settings

Maïke Scherrer, ZHAW School of Engineering, Winterthur, Switzerland, Joel Weingart

The densification of urban settings will further increase in the near future. The already overcrowded roads will no longer be able to absorb the resulting increase in passenger and freight traffic, which will lead to additional congestion. The study at hand analysis if autonomous driving freight vehicles have the potential to defuse the situation.

We analyse the effect of autonomous driving freight vehicles, which commute between a central warehouse and different city districts, where the freight is handed over to cargobikes for the very last mile. We compare this to traditional direct deliveries using vans and cargobikes.

The analysis bases on a route planning and optimisation model, which is specifically adapted for the different vehicle categories and is tested with real data from Switzerland in two settings: First, in a small and a big city, and second, in the case where a central warehouse is close to the city centre compared to a central warehouse in the outskirt of a city.

The results show that higher traffic reduction potential is given in bigger than in smaller cities and in cases where the central warehouse is in the outskirt of the city.

The results can be used in several ways: First, the model can be used to conduct similar analyses in different urban settings. Second, the results can be used to decide about the modal mix for urban logistics and in which situations a change from traditional deliveries to deliveries with autonomous driving freight vehicles in combination with

cargobikes is beneficial.

2 - Incremental Approximate Single-Source Shortest Paths with Predictions

Aidin Niaparast, Carnegie Mellon University, Pittsburgh, PA, United States, Samuel McCauley, Benjamin Moseley, Helia Niaparast, Shikha Singh

The algorithms-with-predictions framework has been used extensively to develop online algorithms with improved beyond-worst-case competitive ratios. Recently, there is growing interest in leveraging predictions for designing data structures with improved beyond-worst-case running times.

In this paper, we study the fundamental data structure problem of maintaining approximate shortest paths in incremental graphs in the algorithms-with-predictions model. Given a sequence of edges that are inserted one at a time, the goal is to maintain approximate shortest paths from the source to each vertex in the graph at each time step. Before any edges arrive, the data structure is given a prediction of the online edge sequence which is used to "warm start" its state.

As our main result, we design a learned algorithm that maintains $(1+\epsilon)$ -approximate single-source shortest paths, which runs in $O(m \eta \log(W)/\epsilon) \cdot \text{polylog}(n)$ time, where W is the weight of the heaviest edge and η is the prediction error.

Our algorithms are consistent (performing nearly as fast as the offline algorithm) when predictions are nearly perfect, have a smooth degradation in performance with respect to the prediction error and, in the worst case, match the best offline algorithm up to logarithmic factors. That is, the algorithms are "ideal" in the algorithms-with-predictions model.

As a building block, we study the offline incremental approximate single-source shortest-path (SSSP) problem, where the edge sequence is known a priori. Note that the offline problem is defined in the worst-case setting (without predictions) and is of independent interest.

3 - Milp-Based Optimization of Passenger Routing in Urban Air Mobility Systems

DongKyun Kim, Department of Battery-Smart Factory, Korea University, Seongbuk-gu, Korea, Republic of, Taesu Cheong, Min Jei Park

Urban Air Mobility (UAM) is an emerging transportation paradigm that utilizes electric vertical take-off and landing (eVTOL) aircraft to mitigate urban congestion and enhance mobility in dense metropolitan areas. Effective planning and coordination are essential to ensure efficient routing, high service quality, and optimal use of limited aerial infrastructure. This study introduces a Mixed-Integer Linear Programming (MILP) model for passenger routing in a UAM network consisting of vertiports, vertistops, and a constrained fleet of eVTOL aircraft. The formulation jointly optimizes routing and scheduling decisions while accounting for critical operational constraints, including vehicle capacity, time window adherence, selection of charging nodes, and airspace connectivity. The objective function minimizes total operational cost, combining flight distances with passenger delay penalties. Flow variables and binary decision variables are used to model feasible, time-consistent flight paths for each passenger, ensuring that vehicle operations and passenger assignments remain synchronized across the network. Computational experiments conducted on representative urban scenarios show that the proposed model can efficiently generate feasible and high-quality routing plans. The results demonstrate its capability to support strategic planning and tactical operations in UAM systems under realistic constraints.

4 - Distance-Based Discount Policy Optimization for Mobile Charging Vehicles

Hua Li, Yonsei University, Seoul, Korea, Republic of, Bongju Jeong

As cities scale up electric vehicle (EV) adoption, mobile charging vehicles (MCVs) offer a promising alternative to fixed charging infrastructure. This study investigates how tiered, distance-based pricing affects user assignment decisions and operator profit in on-demand MCV services. We propose a discrete pricing scheme where users are charged based on distance intervals to available MCVs, capturing spatial willingness-to-pay heterogeneity. Using a simplified assignment model under fixed MCV locations, we simulate service outcomes under various pricing structures and user distributions. Results reveal a non-trivial trade-off between revenue maximization and service accessibility. Strategic segmentation of distance-based fees leads to increased assignment efficiency and higher total profit compared to uniform pricing. The findings highlight the importance of spatial pricing design in improving both customer coverage and economic performance of mobile EV charging systems. This work provides practical insight for operators aiming to deploy MCVs in urban areas with heterogeneous demand profiles.

WB26

Building A Level 4 A412

Baysian Optimization for Complex Systems

Invited Session

Simulation

Chair: Giulia Pedrielli, Arizona State University, Tempe, AZ, United States

Co-Chair: Antonio Candelieri, University of Milano-Bicocca, Milano, Italy

Co-Chair: Raul Astudillo, California Institute of Technology, Pasadena, CA, United States

1 - Wasserstein Barycenters of Gaussian Processes for Exotic Bayesian Optimization Tasks

Antonio Candelieri, University of Milano-Bicocca, Milano, Italy

Given a set of probability distributions, the Wasserstein Barycenter is a new distribution minimizing the *weighted* Wasserstein distance from those in the original set. The Wasserstein Barycenter shows important properties, such as *shape preservation* and (intrinsic) *geometry awareness*, leading this notion to be at the core of many applications in statistics, mathematics and, more recently, Machine/Deep Learning and Generative Artificial Intelligence.

Exploiting the analogy between Gaussian Processes (GPs) and Gaussian Distributions (GDs) -- for which Wasserstein distance and Barycenter show important properties - the notion of weighted Wasserstein Barycenter is applied into the Bayesian Optimization (BO) framework to naturally deal with *exotic* tasks, specifically: federated/collaborative BO, batch-BO with adaptive batch's size, and multi-fidelity / multiple information source BO.

To achieve these goals, weights of the Wasserstein Barycenter have to be suitably identified depending on the specific application, possibly leading to multiple barycenters (e.g., in batch-BO). Indeed, although assuming equal weights represent a reasonable choice in absence of explicit knowledge about the relevance of each probability distribution in the original set, this choice is not optimal with respect to the specific loss (aka objective) function associated to the target problem, as already reported in the literature.

Empirical results on a set of test problems will be discussed along with theoretical considerations about the impact of the Wasserstein Barycenters of GPs on some of the most widely used acquisition functions.

2 - Clad-Net: Continual Activity Recognition in Multi-Sensor Wearable Systems

Reza Rahimi Azghan, Arizona State University, Tempe, AZ, United States, Gautham Krishna Gudur, Mohit Malu, Edison Thomaz, Giulia Pedrielli, Pavan Turaga, Hassan Ghasemzadeh

The rise of deep learning has significantly advanced human behavior monitoring using wearable sensor data. In particular, human activity recognition (HAR) through deep models has been extensively explored. However, most existing methods assume a stationary data distribution, an assumption that often breaks down in real-world settings. For instance, sensor data collected from one subject performing an activity may differ substantially in distribution from that of another subject. In continual learning settings, this shift is modeled as a sequence of tasks—where each task corresponds to a new subject—which makes the system susceptible to catastrophic forgetting, where previously learned patterns degrade as new ones are acquired. This challenge is further exacerbated by the reliance on labeled data, which is often sparse or inconsistently available in studies involving human participants. To address these challenges, based on the idea of complementary learning systems, we introduce CLAD-Net, Continual Learning with Attention and Distillation. Our model combines a self-supervised transformer, serving as the system's long-term memory, with a supervised CNN trained with knowledge distillation for activity classification. The transformer captures global activity patterns through cross-attention across body-mounted sensors and learns generalizable representations without requiring labels. In parallel, the CNN incorporates knowledge distillation to retain past knowledge during subject-wise

fine-tuning. We evaluate CLAD-Net under a continual stream of data from multiple subjects and demonstrate its strong resilience to forgetting. We also conduct experiments to study the individual contributions of the different components used in each of the models.

3 - Conditional Diffusion Models for Bayesian Optimization

Szu Hui Ng, National University of Singapore, Singapore, Singapore, Yilin Zheng, Haowei Wang

Bayesian Optimization (BO) is a sample-efficient framework for optimizing expensive black-box functions, typically using a Gaussian Process (GP) as a surrogate model and an acquisition function to guide the sequential search. Among various acquisition functions in BO, information-based methods such as Predictive Entropy Search (PES) aim to reduce uncertainty on the location of the global optimum x^* , but often incur high computational cost due to the difficulty of sampling x^* from the GP posterior. To address this issue, we propose a novel framework that integrates diffusion models into BO, enabling efficient approximation of the distribution over the global optimum x^* . Within this framework, we first design a standardized suite of training guidelines for the diffusion models. With the diffusion model, we propose Diffusion-based PES (DPES) algorithm that significantly reduce computational cost. We further introduce Diffusion-based Mode Seeking (DMS), a new acquisition function that directly selects the mode of the learned distribution as the next query, offering a simpler and more efficient alternative to entropy-based strategies. Empirical results on benchmark functions demonstrate the effectiveness of our proposed approaches.

4 - Mixed Pareto Set Learning for Mixed Variable Multi-Objective Bayesian Optimization

Zijian Huang, University of Michigan, Ann Arbor, MI, United States, Weishi Wang, Raed Al Kontar

Real-world optimization often involves simultaneously addressing multiple expensive black-box objectives defined over mixed discrete and continuous variables, known as Mixed Variable Multi-Objective Bayesian Optimization (MVMBO). This context presents notable computational and modeling challenges, particularly due to the non-smooth and potentially infinite mixed Pareto set. To overcome these difficulties, we introduce Mixed Pareto Set Learning (MPSL), a novel framework designed for efficiently recovering the mixed Pareto set within limited evaluation budgets. MPSL employs a relaxed mapping strategy that translates decision-makers' trade-off preferences into Pareto-optimal solutions via probabilistic parameterization of discrete variables, effectively navigating the complexity inherent in mixed design spaces. Additionally, we propose a model-agnostic local search refinement procedure, empirically shown to enhance candidate solution quality and hypervolume performance significantly. Our refinement method demonstrates clear advantages over simply increasing preference sampling and exhibits versatility by benefiting general Multi-Objective Bayesian Optimization (MOBO) scenarios beyond MVMBO. Extensive numerical experiments, including synthetic benchmarks and a real-world electrohydrodynamic (EHD) inkjet printing application, validate that MPSL offers superior Pareto front coverage, improved computational efficiency, and real-time flexibility for decision-makers, underscoring its practical importance across critical engineering and scientific domains.

5 - Bayesian Experimental Design with Generative Input Priors: a KI-Regularized Decision-Theoretic Framework

Raul Astudillo, California Institute of Technology, Pasadena, CA, United States

Bayesian experimental design (BED) is a powerful framework for data-efficient learning and optimization. Yet, standard methods often overlook structural prior knowledge about the input space—such as feasibility constraints or low-dimensional manifolds—that can now be effectively captured by generative models. We introduce a KL-regularized decision-theoretic framework that integrates such generative priors into BED. This formulation yields a simple, principled characterization of one-step Bayes-optimal policies as Boltzmann distributions over the input space, naturally balancing task-specific utility with fidelity to the learned prior. Our perspective unifies and generalizes previous approaches that incorporate generative models into Bayesian optimization, and it lays the groundwork for more advanced developments, including multi-step design strategies and theoretical regret guarantees.

WB41

Building B Level 3 B301

AI for Smart Health: Modeling, Simulation, and Optimization

Invited Session

Health Applications Society

Chair: Bing Yao, The University of Tennessee Knoxville, Knoxville, TN, United States

1 - Simulation Optimization of Spatiotemporal Dynamics in 3D Geometries

Bing Yao, The University of Tennessee Knoxville, Knoxville, TN, United States, Fabio Leonelli, Hui Yang

Many engineering and healthcare systems are featured with spatiotemporal dynamic processes. The optimal control of such systems often involves sequential decision making. However, traditional sequential decision-making methods are not applicable to optimize dynamic systems that involves complex 3D geometries. Simulation modeling offers an unprecedented opportunity to evaluate alternative decision options and search for the optimal plan. In this paper, we develop a novel simulation optimization framework for sequential optimization of 3D dynamic systems. We first propose to measure the similarity between functional simulation outputs using coherence to assess the effectiveness of decision actions. Second, we develop a novel Gaussian Process (GP) model by constructing a valid kernel based on Hausdorff distance to estimate the coherence for different decision paths. Finally, we devise a new Monte Carlo Tree Search (MCTS) algorithm, i.e., Normal-Gamma GP MCTS (NG-GP-MCTS), to sequentially optimize the spatiotemporal dynamics. We implement the NG-GP-MCTS algorithm to design an optimal ablation path for restoring normal sinus rhythm (NSR) from atrial fibrillation (AF). We evaluate the performance of NG-GP-MCTS with spatiotemporal cardiac simulation in a 3D atrial geometry. Computer experiments show that our algorithm is highly promising for designing effective sequential procedures to optimize spatiotemporal dynamics in complex geometries.

2 - Robust Coupled Tensor Decomposition for Shared and Personalized Feature Selection Under Multiple Conditions

Ana Estrada Gomez, Purdue University, West Lafayette, IN, United States, Manni Zhang

Nowadays, complex systems are monitored by hundreds of sensors. How to extract useful information from the multi-modal data for system diagnosis and control is still an open question. In this work, we aim to identify the most relevant information or features collected by the sensors for system diagnosis and control. Different challenges need to be addressed. (1) Systems operate under different conditions. It is critical to differentiate between condition-specific features and system-specific features to account for data heterogeneity. (2) High-dimensional data from different sensors needs to be fused. Features within a data modality offer a partial perspective of the system, whereas fused data yields a comprehensive view of the system. (3) It is common for sensor data to be missing or corrupted. It is critical to develop a robust method for fault isolation and identification. To tackle these challenges, we propose a framework, Missing-value-imputed, RObust, COupled, Sparse and Similar CANDECOMP/PARAFAC Decomposition (MissROCOSS-CPD), that fuses multi-modal data and exploits the similarities and differences across conditions for feature selection. Our method decomposes each modality into a low-rank component that captures the true signal and a sparse component that captures the outliers. We simultaneously learn the joint and individual factor matrices across modalities from the low-rank components and identify the shared and personalized features by adding sparse group Lasso penalties to the individual factor matrices. The performance of the proposed framework is illustrated through simulated datasets and a case study.

3 - Decoding Emotion from Eeg Using Weighted Heterogeneous Recurrence Network Analysis

Yujie Wang, University of Miami, Miami, FL, United States, Cheng-Bang Chen

Emotion recognition from electroencephalography (EEG) signals remains a significant challenge due to the brain's nonlinear, dynamic, and nonstationary activity. This work proposes a novel framework that combines manifold learning, Weighted Heterogeneous Recurrence Network Analysis (WHRNA), and ensemble classification to capture and interpret emotional dynamics in EEG data. High-dimensional EEG signals are first embedded into a low-dimensional manifold using Uniform Manifold Approximation and Projection (UMAP), preserving spatial structure while enhancing computational efficiency. These embedded representations form temporal trajectories, which are segmented via Voronoi tessellation to account for heterogeneous recurrence behaviors across time and space. WHRNA then models multiscale nonlinear dynamics by constructing weighted recurrence networks from these segments. Finally, tree-based ensemble models classify emotional states with an improved accuracy, which achieves 83.7% on SEED and 79.1% on SEED-IV datasets, surpassing prior benchmarks. Sensitivity analyses reveal specific WHRNA-derived features

associated with distinct emotions, offering insights into the neural substrates of affective states. This approach enhances both performance and interpretability, providing a powerful tool for affective computing, mental health assessment, and cognitive neuroscience.

WB42

Building B Level 3 B302

Data-driven Policy Making for Healthcare and Social Responsibility

Invited Session

Health Applications Society

Chair: Shinyi Wu, University of Southern California, Santa Monica, CA, United States

Co-Chair: Jiyang He, University of Southern California, Los Angeles, CA, United States

1 - Improving Estimation of Pediatric Mental Health Prevalence in Washington State Under a Marginal Distributions Constrained Optimization Framework

Shan Liu, University of Washington, Seattle, WA, United States, Yinsheng Wang, Kathryn Brandeau

Pediatric mental health disorders, such as depression, anxiety, and behavioral disorders, significantly impact the well-being of children and youth. In Washington state (WA), more than one-third of the population resides in regions lacking adequate mental health providers. Accurately mapping the geographic distribution of mental health demand is thus essential for designing optimal care networks for pediatric services. This study estimates the prevalence of pediatric mental health conditions by integrating multiple data sources, including the National Survey of Children's Health (NSCH), American Community Survey (ACS), and local health surveys and hospital discharge data from WA. Initially, we developed machine learning regression models using NSCH data, incorporating key demographic factors such as age, race, sex, family structure, and poverty status. Subsequently, we constructed a marginal distributions constrained optimization model to recover detailed demographic group proportions by harmonizing regression estimates from the NSCH with corresponding estimates derived from multiple local datasets, constrained by demographic marginal distributions obtained from the ACS. Analysis of surveys and hospital data from WA enabled robust local calibration of prevalence ratios. This integrative modeling approach enhances small-area prevalence estimation by effectively fusing national and local datasets, facilitating precise resource allocation and informing strategic placement decisions for pediatric mental health facilities and programs.

2 - Informing Effective, Efficient, and Equitable Harm Reduction Interventions Through Decision Modeling: Lessons from the ProFOUND/Rescue Project

Xiao Zang, University of Minnesota, Minneapolis, MN, United States

The ongoing overdose crisis in the United States demands urgent, data-driven responses that are not only effective and cost-efficient, but also equitable. This presentation highlights the use of simulation modeling, decision analysis, and optimization methods to inform harm reduction strategies, drawing from the NIH/NIDA-funded PROFOUND/RESCUE project. By integrating epidemiologic and health services data, this project evaluates and optimizes the allocation of naloxone and other harm reduction interventions across diverse populations and geographies in six U.S. jurisdictions.

We developed an individual-based microsimulation model and accompanied decision support tools, including interactive dashboards and Shiny Apps, to assess a wide range of intervention scenarios, varying in geographic targeting, distribution channels, and population reach. Calibrated to local epidemiological data, the model estimates outcomes such as overdose deaths averted, costs, and equity impacts. We also developed novel analytical frameworks to define and measure naloxone availability and saturation, and to optimize its geographic distribution. These frameworks embed simulation-based analysis to identify strategies that maximize public health impact within constrained resources, with extensions that directly incorporate equity considerations into decision-making.

This work illustrates the power of simulation and optimization to guide evidence-based policy decisions in complex and resource-limited environments. The presentation will highlight key insights, methodological innovations, and future research opportunities at the intersection of decision science, optimization, and public health policy.

3 - Building Ethical AI/ML Decision Making Systems for Justice Reform

Aparna Komarla, Redo.io, Berkeley, CA, United States

In 2018, California became the first state in the US to pass a Prosecutor-Initiated Resentencing law, which grants prosecutors the [power](#) to revisit past convictions. In addition, a number of resentencing and second-look laws in the state, such as AB600 and the Racial Justice Act, seek to correct past injustices in the criminal legal system. But given that there are nearly 100,000 people in California's prisons, lawyers and county officials need efficient and accurate ways to identify cases worth a second-look. We estimate that there are hundreds languishing in California's prisons as a result of technological and political barriers. Redo.io (www.redoio.info) is solving this problem with predictive, exploratory and rules-based models for resentencing eligibility determination. Our open-source AI assistant helps attorneys quickly query our dataset and find prison sentences that align with their criteria. Our predictive models embody the principles of fairness, accountability and transparency and help prioritize cases based on violent history and commitment to rehabilitation (education, mental health treatment). Legal experts have helped design and fine-tune eligibility rules in an iterative process. Our model, by design, augments legal knowledge about the applicability of second-look laws as opposed to replacing such expertise. By supplementing a lawyer's decision to examine a case for resentencing instead of automating the end-to-end process, we help stakeholders make high-impact decisions efficiently while preventing existing biases in our society from being exacerbated. We therefore exemplify a framework for meaningful human-computer interactions in high-stakes and high-risk decision making processes.

4 - Adapting Behavioral Health Models Across Policy Shifts: A Transfer Learning Approach to Cannabis Dependence Prediction

Jiyang He, University of Southern California, Los Angeles, CA, United States, Shinyi Wu, Carolyn Wong

Rapid policy changes in cannabis regulation, combined with evolving user demographics, create significant challenges for behavioral health risk models. Patterns learned from older pre-legalization populations may not generalize to younger post-legalization cohorts due to regulatory, market and generational shifts. This study addresses the critical need for adaptive modeling approaches that can efficiently update risk prediction models across these transitions.

We developed a transfer learning (TL) framework to predict problematic cannabis use (PCU) across three distinct contexts using longitudinal data from the CHAYA study (N=506). Our approach employs LSTM networks in a pretrain-fine-tune architecture, transferring knowledge from pre-legalization older cohorts (2014-2018, ages 18-26 at baseline) to post-legalization younger populations (2019-2023, ages 18-21 at baseline). We evaluated model performance across varying amounts of target domain data (0%-100%) and conducted interpretability analysis using SHAP values.

Results demonstrate that direct zero-shot transfer fails dramatically when both policy contexts and population demographics shift (AUC dropping from 0.65 to 0.44). However, fine-tuning with just 25% of new target data rapidly restored performance, achieving AUCs of 0.89-0.92 across transfer scenarios. TL outperformed both training from scratch and pooled modeling approaches. Feature importance analysis revealed significant shifts in risk factors, with relationship status at baseline data collection emerging to show strong association with PCU in post-legalization younger cohorts, while mental health remained as a strong indicator stably across domains.

This work demonstrates that TL enables cost-effective adaptation of behavioral health models to changing contexts, providing a practical approach for ensuring model relevance across market, generational and regulatory shifts.

WB44

Building B Level 3 B304

Supply Chain and Logistics Operations

Contributed Session

Contributed

Chair: Asef Sajjad Abir, 515 30th Ave N, Fargo, ND 58102, USA, Fargo, ND, 58102, United States

1 - A Model for Managing Supply Chains with Capacity Investment and Financial Hedging Under Risk from Natural Disasters, Political Conflict**Ismail Civelek, WKU, Bowling Green, KY, United States**

Supply chain management is defined as the efficient integration of manufacturers, suppliers, warehouses, vendors, and customers to maximize value across-the-board by minimizing system wide costs while satisfying customer and fill-rate requirements. Considering recent dramatic disruptions in managing supply chains (i.e., COVID-19, natural disasters, Russian invasion of Ukraine, etc.), supply chain managers must factor in factory shutdowns due to pandemic, conflict, port shutdowns or natural disasters (earthquakes, tsunami, tornadoes, forest fires, etc.). This study proposes an analytical model for a manufacturer, which is under risk due to natural disaster or political risk. A petroleum refinery is a good example because its production mostly depends on the input material, crude oil. A food (or agricultural) processing company or a power plant that uses oil or natural gas would be other motivating examples for this supply chain problem. In addition to related industries, an automotive manufacturer that uses steel, aluminum, and rubber for major material inputs would be another motivating example. The manufacturing company has two joint problems: capacity investment and financial trading. This study models the joint problem and characterizes the optimal policy for single period and finite multi-period settings.

2 - Min-TIME Coverage in Constricted Environments with Heterogeneous Robotic Fleets**Young In Kim, Georgia Institute of Technology, Atlanta, GA, United States, Spiridon Reveliotis**

In a recent research program, we have investigated robotic traffic management problems that arise when a fleet of networked mobile robots is employed to support certain coverage tasks taking place in physically constricted environments.

Our previous studies focused on the case where the employed robots have the same role and capability in the executed task.

In this work, we extend our investigation to the case where robots have distinct roles and capabilities.

We systematically introduce these extensions and formulate the corresponding traffic management problems as integer programs (IPs).

Furthermore, we develop a strong combinatorial relaxation of these IP formulations that significantly reduces the number of integer variables involved without compromising the optimality of the solutions.

Lastly, numerical results demonstrate the effectiveness and the efficiency of the relaxation.

3 - Selling and Leasing Strategies for Perfectly Complementary Products**Wanfen Yang, University of California, Irvine, Irvine, CA, United States, Shuya Yin**

In various industries, we observe a rising trend of hybrid acquisition models in which consumers purchase part of a product bundle while leasing the other part. A prominent example is NIO's Battery-as-a-Service (BaaS) in the electric vehicle industry, where the car body is purchased while the battery is leased. In this paper, we examine a setting involving durable complementary products to explore optimal selling and leasing strategies.

4 - Optimal Replenishment Policy from Dual Supplier with Nonconsecutive Lead TIMES**Chaehoon Han, Seoul National University, Seoul, Korea, Republic of, Kunsoo Park, Byeongkwon Lee**

We study an inventory problem involving dual sourcing from two suppliers with different lead times. While the optimal inventory policy for this problem has been proved only for the case where the lead times of two suppliers are consecutive (i.e. the lead time difference between them is one), this study is, to our best knowledge, the first that proves the optimal replenishment policy with nonconsecutive lead times. We formulate the expected cost function using dynamic programming and verify its convexity through partial optimization and gradient analysis. Then, we demonstrate that the optimal replenishment from the expedited supplier follows the 'pipeline-dependent base-stock policy', where the base-stock level depends on the quantities ordered but not yet received from the regular supplier. Our findings not only address a research gap in dual sourcing inventory system but also introduce a new analytical approach applicable to multi-dimensional inventory problems such as the lost-sales inventory model.

5 - Finding the Optimal Takt Time Group for Individualized Products in Mixed-Model Assemblies**Philipp Pithan, WHU - Otto Beisheim School of Management, Vallendar, Germany, Arnd Huchzermeier**

While assembly lines face increasing pressure to produce locally, customers expect individualized products. These developments are accelerating the shift toward local, mixed-model assembly systems designed to produce highly individualized products — posing new and complex manufacturing challenges. Manufacturers need to balance, schedule, and sequence these products in an efficient yet flexible manner — a impossible challenge with a single fixed takt.

We introduce T-Bars as a simple framework for capturing the variance introduced by selective and additive options and demonstrate how they can be used together with permissible overwork and a variable takt to achieve a perfectly balanced assembly line for multiple individualized products. We analyze the benefits of a variable takt and evaluate the implications of grouping either multiple products or multiple configurations into takt time groups (TTG). The developed mathematical model groups either (i) products or (ii) configurations and selects the best takt time for each group, thus optimizing line utilization.

In a real-world application, we find that the T-Bar framework greatly simplifies the assembly line balancing effort. The use of multiple TTG allows for higher utilization, regardless of whether product- or configuration-based. We show that both have distinct advantages and trade-offs. The former leads to more stable plans, as task-station assignments remain consistent within each product, though lower utilization. The latter allows for higher utilization but results in less stable balancing plans due to frequent changes in task-station allocation. In both cases, the number of TTG should be limited to strike a balance between efficiency and practicability.

6 - Zero-Shot Foundation Model Pipeline for Image Analysis in Biomanufacturing**Rui Qi Chen, Georgia Institute of Technology, Atlanta, GA, United States, Benjamin Joffe, Caroline Serafini, Francisco Robles, Jing Li**

Effective Process Analytical Technology (PAT) in cell therapy manufacturing requires process monitoring and feedback control to ensure product quality. Automated image analysis, leveraging real-time data from non-invasive techniques such as quantitative Oblique Back-illumination Microscopy (qOBM), is integral to achieving this by providing key bioprocess insights. However, current automated image analysis methods face limitations: supervised models demand extensive annotation and struggle with batch effects, while common anomaly detection methods often lack interpretability. Furthermore, the scarcity of comprehensive, well-annotated datasets makes training capable models challenging.

To address these challenges, we propose a novel two-stage pipeline for analyzing qOBM images. First, the Segment Anything Model (SAM) performs zero-shot instance segmentation. Second, a Vision Language Model (VLM) or a multimodal Large Language Model (LLM) with vision capabilities, such as Gemini 2.5 Pro, analyzes these objects within the full image context. This approach leverages pre-trained foundation models to eliminate annotation and training burdens. It enables flexible analysis using prompt-based inputs that embed domain knowledge without requiring model retraining. The broad visual and language understanding of these models improves generalizability and robustness against batch effects and domain shifts. Furthermore, it enhances interpretability by generating natural language explanations and descriptive outputs for detected anomalies. This analytical framework, demonstrated on T-cell cultures, facilitates critical tasks such as real-time cell detection, segmentation, counting, and live/dead classification for viability analysis, alongside explicit identification and description of process deviations like biological contamination. This pipeline is anticipated to advance automated process control and PAT, enabling more reliable, data-efficient biomanufacturing.

7 - The SME Business Maturity Model (SBMM): A Framework for Navigating Growth and Enhancing Operational Effectiveness**Asef Sajjad Abir, North Dakota State University, Fargo, ND, United States, Diana López-Soto**

Small to Medium-sized Enterprises (SMEs), defined as firms with fewer than 249 employees, play a crucial role in the economy, accounting for roughly 95% of global businesses and more than 60% of employment. The global small business market, valued at USD 2,572 billion in 2023, is anticipated to grow to USD 4,985 billion by 2032. Despite its significance, 67% of SME executives identify survival and expansion as key challenges in the face of economic uncertainties, facing difficulties in

transitioning from unstructured operations to sustainable maturity due to insufficient strategic support and funding. This study presents the SME Business Maturity Model (SBMM), a framework that offers a systematic approach for the development of small and medium enterprises (SMEs). The SBMM, based on systematic reviews of SME growth challenges and foundational maturity models, supports advancement from initial stages to established industry leadership. This assessment evaluates critical dimensions such as business model, financial management, operations, technology utilization, and customer acquisition, categorizing SMEs into one of six distinct developmental phases. Each phase exhibits distinct characteristics, challenges, and customized development strategies. The SBMM seeks to systematically improve operational efficiency by identifying areas of attention specific to each phase and recommending actionable strategies. This methodology enables SMEs to develop strong capabilities, enhancing profitability, resilience, and promoting sustainable growth necessary for their survival and expansion.

WB46

Building B Level 3 B306

Online Real-time Intelligent Decision Making Methods and Analysis

Invited Session

Decision Analysis Society

Chair: Mengfan Xu, UMass Amherst, Amherst, MA, United States

1 - Simulation Optimization for Contextually Aware Online Decision Making

Gregory Keslin, New Jersey Institute of Technology, Newark, NJ, United States

Ranking and Selection is a simulation optimization methodology used to solve complex decision problems where each decision is associated with a simulation system. Historically, such problems are static pre-specified problems with a single solution. However, there has been a recent focus on dynamic real-time decision-making problems that utilize real-world contextual information to reduce uncertainty for a better-informed decision-making process. Because classical Ranking and Selection procedures require every decision to be repeatedly simulated, Ranking and Selection procedures are typically ill suited to solve these dynamic problems rapidly or in real time. In this talk, we present a novel approach that runs the entire simulation budget “offline”, prior to observation of the contextual information. A classifier is then fit on the simulation outputs to determine the optimal “online” decision once the contextual information is observed. Associated with the classifier is an assessment of the classifier’s performance. This assessment achieves a finite-sample coverage guarantee under mild conditions.

2 - Sequential Stochastic Combinatorial Optimization Using Hierarchical Reinforcement Learning

Haipeng Chen, William & Mary, Williamsburg, VA, United States, Xinsong Feng, Zihan Yu, Yanhai Xiong

Reinforcement learning (RL) has emerged as a promising tool for combinatorial optimization (CO) problems due to its ability to learn fast, effective, and generalizable solutions. Nonetheless, existing works mostly focus on one-shot deterministic CO, while sequential stochastic CO (SSCO) has rarely been studied despite its broad applications such as adaptive influence maximization (IM) and infectious disease intervention. In this paper, we study the SSCO problem where we first decide the budget (e.g., number of seed nodes in adaptive IM) allocation for all time steps, and then select a set of nodes for each time step. The few existing studies on SSCO simplify the problems by assuming a uniformly distributed budget allocation over the time horizon, yielding suboptimal solutions. We propose a generic hierarchical RL (HRL) framework called wake-sleep option (WS-option), a two-layer option-based framework that simultaneously decides adaptive budget allocation on the higher layer and node selection on the lower layer. WS-option starts with a coherent formulation of the two-layer Markov decision processes (MDPs), capturing the interdependencies between the two layers of decisions. Building on this, WS-option employs several innovative designs to balance the model’s training stability and computational efficiency, preventing the vicious cyclic interference issue between the two layers. Empirical results show that WS-option exhibits significantly improved effectiveness and generalizability compared to traditional methods. Moreover, the learned model can be generalized to larger graphs, which significantly reduces the overhead of computational resources. Our code is released at: <https://github.com/wmd3i/HRL4SSCO>

3 - Probabilistic Non-Negative Factorization with Latent Gaussian Processes

Jialin Li, University of Massachusetts, Amherst, Amherst, MA, United States, Lulu Kang

Developing statistical models for prediction and uncertainty quantification can be unprincipled when we have a priori knowledge about the system (e.g., output non-negativity). In this work, we consider a general modeling problem using expensive observational multi-output data, with recorded configuration/environment parameters for each data point. Within an integrated probabilistic framework, we factorize the non-negative parameter-to-multi-output function into multiple components expressed as “softly censored” latent Gaussian processes. The model is delicately hyperparametrized and we propose a specific variational inference approach to learn the components and hyperparameters. With improved model flexibility and expressive power, we can perform better predictions and correct uncertainty quantification for outputs associated with parameters not yet observed, and the method is compatible with complicated kernel designs.

4 - Asynchronous Decision-Making in Multi-Agent Bandits

Xuchuang Wang, UMass Amherst, Amherst, MA, United States, Yu-Zhen Janice Chen, Xutong Liu, Lin Yang, Mohammad Hajiesmaili, Don Towsley, John C.S. Lui

We study the cooperative asynchronous multi-agent multi-armed bandits problem, where each agent’s active (arm pulling) decision rounds are asynchronous. That is, in each round, only a subset of agents is active to pull arms, and this subset is unknown and time-varying. We consider two models of multi-agent cooperation, fully distributed and leader-coordinated, and propose algorithms for both models that attain near-optimal regret and communications bounds, both of which are almost as good as their synchronous counterparts. The fully distributed algorithm relies on a novel communication policy consisting of accuracy adaptive and on-demand components, and successive arm elimination for decision-making. For leader-coordinated algorithms, a single leader explores arms and recommends them to other agents (followers) to exploit. As agents’ active rounds are unknown, a competent leader must be chosen dynamically. We propose a variant of the Tsallis-INF algorithm with low switches to choose such a leader sequence. Lastly, we report numerical simulations of our new asynchronous algorithms with other known baselines.

5 - Multi-Agent Multi-Armed Bandit with Fully Heavy-Tailed Dynamics

Mengfan Xu, UMass Amherst, Amherst, MA, United States, Xingyu Wang

We study decentralized multi-agent multi-armed bandits, where clients communicate over sparse random graphs with heavy-tailed degree distributions and observe heavy-tailed reward distributions with potentially infinite variance. We are the first to address such fully heavy-tailed scenarios, capturing the dynamics and challenges of communication and inference among multiple clients in real-world systems, and provide regret bounds that match or improve upon existing results developed in even simpler settings. Under homogeneous rewards, we exploit hub-like structures unique to heavy-tailed graphs to aggregate rewards and reduce noises when constructing UCB indices; under M clients and degree distributions with power-law index $\alpha > 1$, we attain a regret (almost) of order $O(M^{1-\frac{1}{\alpha}} \log T)$. Under heterogeneous rewards, clients synchronize by communicating with neighbors and aggregating exchanged estimators in UCB indices; by establishing information delay bounds over sparse random graphs, we attain a $O(M \log T)$ regret.

WB47

Building B Level 3 B307

Optimization Topics IV

Contributed Session

Contributed

1 - Multi-Stage Optimization Models for Walmart Private Fleet Driver Scheduling and Tractor Assignment

Loksagar Rudraraju Subramanyam, Walmart Inc, Bentonville, AR, United States, Ian Fraser, KIRAN KUMAR BANDELI, Shitij Gusain, Zabiulla Mohammed, Arun Kannan

Our research focuses on how Walmart optimizes drivers and tractors utilization for hauling loads throughout the network comprising: vendors, distribution centers, consolidation centers, and stores. Each Walmart's Transportation Office (TO) supporting one or multiple DCs is tasked with efficiently managing drivers and tractors to meet the ebbs and flows of the required workload. This raises the need for optimal driver schedules and tractor assignment. The proposed framework consisting of three multi-objective Mixed Integer Linear Programming (MILP) models was built to integrate complex tactical and operational decisions for scheduling over 16,000 drivers across 100 TOs to assign and rotate 12,800 tractors among them, where the components are: (1) a driver scheduler, which generates ideal annual schedules for drivers to cover workload requirements; (2) a driver hiring recommender, which provides ideal schedules for new hire drivers to reduce transportation costs; and (3) a tractor assigner, which assigns tractors to drivers on a weekly basis to haul loads. By leveraging the MILP models, the aforementioned-framework ensures to minimize transportation operational costs, slip-seating (driver switching the tractors in-between the schedules) and maximizes preferred tractor assignment for certain driver types. The Models ability to quickly adapt to changing inputs, such as tractor unavailability, maintenance schedules for tractors, drivers churn, and drivers time off enables Walmart to always steer towards optimal operational conditions. Implementing the suggested framework helps Walmart Transportation yield intangible potential savings with the operational costs by \$9M and maintain the slip-seating below 7% reducing the inefficiencies with driver schedules.

2 - A New Optimization Approach to Multi-Period Planning for A Vertical Farming Supply Chain

SAEED ABDOLMALEKI, Supply Chain & Analytics Department, University of Missouri - St. Louis, Saint Louis, MO, United States, Haitao Li

We consider the tactical-level planning for a vertical farming supply chain, where vertical farming systems using hydroponic and aeroponic growing technologies and artificial lighting are employed to enable year-round crop production in vertically stacked racks under climate-controlled conditions. We develop a new mixed-integer programming (MIP) model to determine decisions related to internal production (growing, harvesting, processing), storage, distribution of multiple produce to markets, and—when necessary—external sourcing from other producers to fulfill market demand. The objective is to maximize total profitability, defined as total revenue minus the costs of setup, energy, labor, storage, procurement, and penalties for unmet demand. Our model and solution provide a data-driven decision-support tool for vertical farms to optimally match their production and supply with fluctuating demand in a sustainable way.

3 - Solving Process Planning and Scheduling Problems Using the Method of Maximum Weighted Independent Set

Kai Sun, University of Maryland, Baltimore County, Baltimore, MD, United States

Process planning and scheduling (PPS) is a critical yet computationally intractable problem in manufacturing systems. It involves determining process plans, and scheduling and allocating limited resources to complete jobs composed of sequenced operations. A common optimization objective is to minimize the makespan, i.e., the total completion time of all jobs. Due to the combinatorial complexity of PPS, traditional approaches such as mixed-integer programming (MIP), metaheuristics, and genetic algorithms often struggle to deliver high-quality solutions within practical runtimes.

To address the dual challenge of high computational cost and solution quality, prior studies have explored graph coloring-based methods aimed at solving the PPS problem more efficiently. However, these formulations are not readily applicable to the generic PPS problem. Therefore, this work proposes a novel graph-based formulation for the generic PPS problem considering the minimum makespan. By representing the PPS problem as an undirected, weighted conflicting graph, the proposed approach leverages the maximum weighted independent set (MWIS) to select the best set of operations and resources for each discrete time slot.

To further improve both solution quality and runtime, we develop an algorithm framework for the exact solutions and improved approximations of the MWIS problem. Validation on real-world PPS applications demonstrates that the graph-based approach consistently yields feasible and optimal/near-optimal solutions with improved runtimes, outperforming a leading commercial MIP solver, i.e., Gurobi. The developed method is generic and can be extended to other resource planning problems, such as wireless sensor network scheduling and timetable scheduling problems.

4 - Advanced Satellite Scheduling Strategy for Snow Water Equivalent (SWE) Monitoring

Hadis Banafsheh, Arizona State University, Tempe, AZ, United States, divya ramachandran, Paul Grogan

Snow plays a critical role in the global hydrological cycle, acting as a natural reservoir that regulates water availability for ecosystems, agriculture, and human consumption. Accurate estimation of Snow Water Equivalent (SWE) is essential for effective water resource management, particularly in regions where seasonal snowmelt influences river flow, reservoir levels, and drought forecasting.

To improve the scientific return of satellite observations for SWE monitoring, we propose a reward-based scheduling framework that integrates publicly owned and operated satellite systems, offering continuous passive microwave coverage, with emerging taskable satellite services, including commercial Synthetic Aperture Radar (SAR) constellations.

This framework combines diverse satellite resources through a structured, data-driven prioritization process, forming the basis for a scalable and adaptive scheduling strategy. At its core is a scientifically grounded reward function that quantifies the value of potential observations based on measurable geophysical features, such as SWE variability, surface temperature, and sensor-specific constraints. These inputs, derived from remote sensing products and environmental monitoring systems, are used to construct geospatial Efficiency layers, which are then used to calculate observation rewards, that dynamically highlight where and when targeted taskable observations will yield the greatest scientific value.

By continuously updating observation priorities in response to evolving environmental conditions and mission needs, the framework enables more responsive and effective deployment of taskable assets. While developed for SWE monitoring, the methodology is broadly applicable across Earth science domains that require agile observation planning and multi-source data integration.

6 - Behavior-Aware Joint Optimization of Fulfillment Network Design and Product Assortment in Substitutable Demand Networks

Ashwin Pothen, Georgia Institute of Technology, Atlanta, GA, United States, Benoit Montreuil

E-commerce fulfillment and distribution networks face complex challenges due to increasingly diverse customer behaviors, high service expectations, and the need for inventory efficiency in a disruption-prone environment. We present a novel simulation-optimization framework that jointly optimizes the selection of fulfillment centers (FCs) and the allocation of substitutable product assortments, explicitly incorporating customer behavioral heterogeneity. Our model integrates two key behavioral dimensions: willingness-to-wait and substitution preferences, each informed by empirical distributions derived from real-world data.

The methodology embeds a multi-agent simulation within a Physical Internet-inspired network design. The optimization layer simultaneously addresses facility location, assortment planning, and fulfillment routing to minimize system-wide inventory and transportation costs, subject to service-level constraints. Substitution utility functions are estimated from historical customer choice data, allowing the model to allocate inventory in a manner that maximizes expected customer satisfaction under stockout conditions.

Applied to a large-scale U.S. distribution network, our preliminary results indicate that the proposed method achieves equivalent levels of demand fulfillment with 20–30% lower inventory compared to conventional approaches. Moreover, we observe that total inventory requirements do not scale with the number of FCs, due to the strategic placement of substitutable assortments and behavior-aware allocation. These findings demonstrate the potential of integrating behavioral operations with physical network optimization to advance resilient and efficient fulfillment strategies.

WB48

Building B Level 3 B308

Optimization and Learning Applications in Healthcare

Invited Session

Health Applications Society

Chair: Yonatan Mintz, University of Wisconsin Madison, Madison, WI, United States

1 - Clinical Trial Enrollment Timing with Information-Sharing

HANWEN LIU, Clemson University, Clemson, SC, United States, Qi Luo, Amin Khademi

Phase II/III clinical trials are critical for evaluating the efficacy and safety of new treatments prior to approval by the Food and Drug Administration (FDA). Recent studies suggest that some patients may strategically delay their enrollment to gain additional information about the trial's progress. To model such behavior, we propose a dynamic game framework incorporating patients' strategic enrollment timing and information sharing, offering nuanced insights into trial enrollment decision-making.

2 - Designing Hospital-at-Home Systems: AI and Optimization for Accessible, Distributed Care

Holly Wiberg, Carnegie Mellon University, Pittsburgh, PA, United States, Peter Zhang, Hao Hao, Shuwei He, Peiyu Wang

As healthcare systems increasingly turn to Hospital-at-Home (HaH) models to expand access and alleviate pressure on facilities, they must simultaneously determine which patients can be safely treated at home and how to deliver timely, cost-effective care outside the hospital. This project presents an AI-enabled framework that jointly makes these clinical and operational decisions---identifying appropriate patients for HaH and designing the staffing and supply strategy needed to serve them.

We leverage machine learning to predict patient suitability for home-based care, and their associated care needs, based on clinical complexity, resource availability, and population characteristics. These predictions feed into a mixed-integer optimization model that determines who to send home and how to meet their care needs. The model considers tradeoffs between aggregation and personalization and how these choices affect costs, access, and responsiveness under operational constraints and uncertainty in patient trajectories.

Rather than separating clinical assignment and logistics planning, our framework treats them as interdependent choices. We evaluate system performance under varied geographic distributions, patient profiles, and capacity limits to identify the conditions under which HaH is most effective. Our results offer actionable insights for designing scalable, equitable HaH programs that adapt to uncertainty while balancing efficiency and patient-centered care.

3 - Prescriptive Optimization for Adaptive Auto-Insurance Pricing with Telematics Data

Qinyang He, University of Wisconsin Madison, Madison, WI, United States, Yonatan Mintz

The automotive insurance industry is undergoing fundamental transformation as telematics and decision science enable sophisticated, personalized risk assessment to replace traditional one-size-fits-all pricing models. We present a novel prescriptive optimization framework for personalizing insurance premiums through a two-stage process: first, tailored machine learning models predict individual driver claim risk and expected amounts using telematics data; second, an optimization problem incorporating driver behavioral patterns and price sensitivity determines personalized discount rates. Our experimental results demonstrate this framework creates a win-win scenario where insurers experience increased revenue while drivers benefit from safety-based incentives and fair pricing. Notably, the approach improves individual outcomes and enhances overall social welfare by promoting safer driving behaviors across the insured population.

4 - A Power Constrained Posterior Sampling Algorithm for Nonstationary Bandits with Habituation and Recovery Dynamics

Fengxu Li, University of Wisconsin-Madison, Madison, WI, United States, Yonatan Mintz

Multiarmed bandits (MABs) are a widely used framework for sequential decision-making under uncertainty, with significant applications in personalized healthcare. While traditional MAB methods typically assume stationary reward distributions, real-world settings are often nonstationary. For instance, in personalized interventions, treatment effectiveness may decline due to habituation and later recover, creating complex nonstationary behaviors that standard methods fail to capture.

In addition to adapting to these evolving reward structures, a critical challenge—especially in clinical trial contexts—is preserving statistical power for reliable inference. Existing bandit algorithms generally optimize cumulative rewards but often neglect adequate exploration which is necessary for detecting meaningful treatment effects.

We propose a novel Posterior Sampling algorithm explicitly designed to tackle this dual challenge. Our method adapts to nonstationary environments characterized by habituation and recovery dynamics, while simultaneously satisfying specified power constraints through controlled exploration.

We provide theoretical guarantees via sublinear regret bounds and validate the approach on two real-world datasets by simulation: a bipolar disorder clinical trial and a microrandomized trial promoting workplace standing. Results demonstrate the method's effectiveness in personalizing interventions while maintaining accurate statistical inference.

WB49

Building B Level 3 B309

Flexibility and Sustainable Operation of Data Centers

Invited Session

ENRE: Energy-Climate

Chair: Grant Ruan, MIT, Boston, MA, United States

1 - Reliability Issues of Grid-Connected Data Centers

Grant Ruan, MIT, Boston, MA, United States

The surge of data center energy use has been broadly recognized, but electric grid operators have limited experience and understanding of the reliability risks. One of the major concerns of infrastructure reliability and fault tolerance is on the sub-synchronous resonance (SSR) incidents triggered by data centers. It remains largely uncertain whether and how the interactions between data centers and the grid may trigger undesirable disturbances and potential stability risks.

Here, I will conduct an in-depth investigation from the grid operator perspective to study the impedance characteristics of data center systems, with a special focus on the magnitude and phase variations within the sub-synchronous frequency range. The sensitivity analysis will be conducted to reveal the most critical control parameters or structural settings in data center systems. Scanning through diverse working conditions, these sensitivity models provide useful information for mitigation strategies such as active impedance shaping. Extensive simulations validate that the proposed model and algorithms successfully identify the sub-synchronous risks and outperform the benchmarks in terms of modeling precision and robustness performances.

2 - Toward Sustainable Data Centers for Artificial Intelligence

Benjamin Lee, University of Pennsylvania, Philadelphia, PA, United States

As the impact of artificial intelligence (AI) continues to proliferate, computer architects must assess and mitigate its energy demands and environmental impact. This talk will survey strategies for mitigating the energy used by AI computation and datacenter infrastructure, drawing on data and experiences from industrial, hyperscale systems. First, we analyze the energy and carbon implications of super-linear AI growth. Second, we re-think datacenter infrastructure and define a solution space for sustainable computation with renewable energy, utility-scale batteries, and job scheduling. Finally, we develop strategies for datacenter demand response, incentivizing workloads to modulate power usage in ways that reflect their performance goals. In summary, the talk provides a broad perspective on sustainable computing and outlines the many remaining directions for future work.

3 - Fine-Grained Carbon Accounting for Serverless Cloud Computing

Prateek Sharma, Indiana University, Bloomington, IN, United States

Serverless computing, or Functions as a Service (FaaS), has emerged as a key cloud abstraction which is enabling rapid deployment of AI, IoT, and scientific computing applications. Cloud functions are ephemeral and small—with millisecond-scale execution times. This necessitates /fine-grained/ accounting and optimization of carbon footprint of individual function invocations.

This talk will present the challenges and recent developments in statistical disaggregation and fair accounting for function-level carbon footprint measurement. For cloud applications, the energy and carbon footprint is diffused across many local and remote software and hardware components. For instance, virtualization and containerization software layers which execute and orchestrate applications (functions in our case) can themselves be significant energy consumers. As a result, existing energy profilers only provide a small and incomplete picture. We show how Shapley value principles can be used for fairly dividing this shared overhead among highly heterogeneous and dynamic functions, improving accuracy significantly. For fine-grained applications like serverless functions, we also develop new validation techniques and metrics, such as marginal energy, which provide a practical approach to obtain the "ground truth" energy and carbon footprint.

4 - Theoretical Foundations of Carbon-Aware Workload Scheduling

Mohammad Hajiesmaili, UMass Amherst, Amherst, MA, United States

As efforts to reduce the environmental impact of computing grow, *carbon-aware workload scheduling* has emerged as a promising approach for improving the sustainability of data centers. In this talk, we introduce a new online decision-making problem inspired by this setting: the *pause-and-resume problem*. Here, the goal is to select the k lowest prices (e.g., carbon intensity levels) from a sequence that is revealed one step at a time. At each time step, a decision-maker must choose whether to accept or reject the current price, while also managing *switching costs* incurred when changing decisions—such as the cost of pausing or resuming a workload.

This problem is motivated by real-world scenarios where computational tasks can be deferred during periods of high carbon intensity and resumed when cleaner energy is available—but doing so incurs operational overhead. The problem shares features with classical models in online optimization but also presents new challenges that render existing solutions insufficient. We propose and analyze a class of *double-threshold algorithms*, extending previous work on threshold-based strategies. We show that these algorithms achieve optimal competitive performance among all deterministic online methods.

This talk will highlight the interplay between sustainability goals, algorithm design, and online decision-making, with implications for broader areas such as carbon-aware and real-time resource management.

WB51

Building B Level 3 B311

Stochastic Programming and Distributionally Robust Optimization Under Decision-Dependent Uncertainty

Invited Session

OPT: Optimization Under Uncertainty

Chair: Miguel Lejeune, George Washington University, Washington, DC, United States

1 - Chance-Constrained Set Multicover Problem

Pavlo Krokhmal, University of Arizona, TUCSON, AZ, United States, Shunyu Yao, Neng Fan

We consider a version of the set covering problem with uncertain parameters, which we refer to as the chance-constrained set multicover problem (CC-SMCP). In this problem, we assume that there is uncertainty regarding whether a selected set can cover an item, and the objective is to determine a minimum-cost combination of sets that covers each item i at least k_i times with a prescribed probability. To tackle CC-SMCP, we construct exact equivalent deterministic reformulations that feature a hierarchy of bounds and develop the corresponding outer-approximation (OA) algorithm. Additionally, we consider reducing the number of chance constraints via vector dominance relations and reformulate special cases of CC-SMCP using the "log-transformation" method and binomial distribution properties. Theoretical results on sampling-based methods, i.e., the sample average approximation (SAA) method and the importance sampling (IS) method, are also developed to approximate the true optimal value of CC-SMCP under a finite discrete probability space. Our numerical experiments demonstrate the effectiveness of the proposed OA algorithm, particularly in scenarios with sparse probability matrices, outperforming sampling-based approaches in most cases and validating the practical applicability of our solution approaches.

2 - Distributionally Robust Chance-Constrained Optimization Under Decision-Dependent Distributional Shift

Miguel Lejeune, George Washington University, Washington, DC, United States

We study distributionally robust chance-constrained optimization problems under decision-dependent uncertainty. The focus is on decision-dependent uncertainty of Type 1 in which the support of the random variables is decision-dependent and shifted. We study various types of coupling functions and ambiguity. We derive computationally tractable equivalent reformulations and a cutting-plane algorithm. Computational tests are reported.

3 - Distributionally Robust Performative Optimization

Zhuangzhuang Jia, University of Illinois Urbana-Champaign, Champaign, IL, United States, Yijie Wang, Roy Dong, Grani Adiwenana Hanasusanto

In performative stochastic optimization, decisions can influence the distribution of random parameters, rendering the data-generating process itself decision-dependent. In practice, decision-makers rarely have access to the true distribution map and must instead rely on imperfect surrogate models, which can lead to severely suboptimal solutions under misspecification. To address this challenge, we propose a distributionally robust framework for performative optimization that explicitly accounts for ambiguity in the decision-dependent distribution. Our framework introduces two modeling paradigms that capture a broad range of applications in machine learning and decision-making under uncertainty. This latter setting has not previously been explored in the performative optimization literature. To tackle the intractability of the resulting nonconvex objectives, we develop an iterative algorithm named repeated robust risk minimization, which alternates between solving a decision-independent distributionally robust optimization problem and updating the ambiguity set based on the previous decision. This decoupling ensures computational tractability at each iteration while enhancing robustness to model uncertainty. We provide reformulations compatible with off-the-shelf solvers and establish theoretical guarantees on convergence and suboptimality. Extensive numerical experiments in strategic classification, revenue management, and portfolio optimization demonstrate significant performance gains over state-of-the-art baselines, highlighting the practical value of our approach.

4 - Stackelberg Games with K-Submodular Function Under Distributional Risk-Receptiveness and Robustness

Sunghoon Park, Virginia Tech, Blacksburg, VA, United States, Manish Bansal

We study submodular optimization in adversarial context, applicable to machine learning problems such as feature selection using data susceptible to uncertainties and attacks. We focus on Stackelberg games between an attacker (or interdictor) and a defender where the attacker aims to minimize the defender's objective of maximizing a k -submodular function. We allow uncertainties arising from the success of attacks and inherent data noise, and address challenges due to incomplete knowledge of the probability distribution of random parameters. Specifically, we introduce Distributionally Risk-Averse k -Submodular Interdiction Problem (DRA k -SIP) and Distributionally Risk-Receptive k -Submodular Interdiction Problem (DRR k -SIP) along with finitely convergent exact algorithms for solving them. The DRA k -SIP solution allows risk-averse interdictor to develop robust strategies for real-world uncertainties. Conversely, DRR k -SIP solution suggests aggressive tactics for attackers, willing to embrace (distributional) risk to inflict maximum damage, identifying critical vulnerable components, which can be used for the defender's defensive strategies. The optimal values derived from both DRA k -SIP and DRR k -SIP offer a confidence interval-like range for the expected value of the defender's objective function,

capturing distributional ambiguity. We conduct computational experiments using instances of feature selection and sensor placement problems, and Wisconsin breast cancer data and synthetic data, respectively.

5 - A Successive Refinement Algorithm for Tri-Level Stochastic Defender-Attacker Problems with Decision-Dependent Probability Distributions

Hugh Medal, University of Tennessee, Knoxville, TN, United States, Samuel Affar

Tri-level defender-attacker game models are a well-studied method for determining how best to protect a system (e.g., a transportation network) from attacks. Existing models assume that defender and attacker actions have a perfect effect, i.e., system components hardened by a defender cannot be destroyed by the attacker, and attacked components always fail. Because of these assumptions, these models produce solutions in which defended components are never attacked, which may not be realistic in some contexts. This study considers an imperfect defender-attacker problem in which defender decisions (e.g., hardening) and attacker decisions (e.g., interdiction) have an imperfect effect such that the probability distribution of a component's capacity depends on the amount of defense and attack resources allocated to the component. Thus, this problem is a stochastic optimization problem with decision-dependent probabilities and is challenging to solve because the deterministic equivalent formulation has many high-degree multilinear terms. To address the challenges in solving this problem, we propose a successive refinement algorithm that dynamically refines the support of the random variables as needed, leveraging the fact that a less-refined support has fewer scenarios and multilinear terms and is, therefore, easier to solve. A comparison of the successive refinement algorithm versus the deterministic equivalent formulation on a tri-level stochastic maximum flow problem indicates that the proposed method solves many more problem instances and is up to 66 times faster. These results indicate that it is now possible to solve tri-level problems with imperfect hardening and attacks.

WB52

Building B Level 3 B312

Stochastic Optimization

Contributed Session

Contributed

Chair: Zhongling Xu, The University of Texas at Austin, Austin, TX, United States

1 - Convergence and Bound Computation for Chance Constrained Distributionally Robust Models Using Sample Approximation

Jiaqi Lei, Northwestern University, Evanston, IL, United States, Sanjay Mehrotra

This paper considers a distributionally robust chance constraint model with a general ambiguity set. We show that a sample based approximation of this model converges under suitable sufficient conditions. We also show that upper and lower bounds on the optimal value of the model can be estimated statistically. Specific ambiguity sets are discussed as examples.

2 - Multi-Stage Stochastic Optimization on Perishable Goods Inventory from Retailers Perspective

Weiliang Deng, University of Arizona, Tucson, AZ, United States

An increasing number of households are struggling with food insecurity, while a significant amount of food is wasted in the supply chain. Multi-parties within the entire supply chain are now collaborating to reduce waste. However, there is a lack of effective inventory management strategies for perishable goods from the retailer's perspective, particularly addressing food waste and donation. This study presents a multi-product, multi-period inventory control model, considering quality- and price-dependent demand and tax incentives to maximize the retailer's overall profit. This model can provide the retailer with an optimal inventory control plan, determining the optimal initial inventory, donation amount, re-order time and re-order amount. Algorithm is designed to solve the problem and numerical experiments are performed to validate the effectiveness of the model.

3 - Mean-Risk Stochastic Mixed-Integer Programs with Endogenous Uncertainty: Deterministic Equivalent Problem Formulations and Decomposition

Lina Villa, Texas A&M University, College station, TX, United States, Lewis Ntaimo

Mean-risk stochastic mixed-integer programs (MR-SMIPs) incorporate risk measures into classical stochastic programming to account for both expected outcomes and variability. Two-stage MR-SMIPs with decision-dependent (endogenous) uncertainty involve here-and-now first-stage decisions that affect the realization of uncertain parameters in the second stage. This work focuses on decision-dependent uncertainty where the first-stage decisions influence the availability of future information. We apply this modeling framework to optimal fuel (vegetation) treatment planning (FTP) under uncertainty, a critical problem in wildfire management. FTP includes methods such as prescribed burning and mechanical thinning to reduce hazardous fuels and thus influence future fire occurrence and behavior, which are both uncertain and driven by weather conditions. We derive two-stage MR-SIP models for the FTP problem with the expected excess (EE) risk measure, conditional value-at-risk (CVaR), and quantile deviation (QDEV), respectively, and create deterministic equivalent problem (DEP) formulations. Due to endogenous uncertainty, the DEP models are nonlinear and challenging to solve. To address this issue, we propose a linearization of the models that preserves the structure and improves tractability. We perform a computational study based on real-life data to evaluate the performance of the DEP models relative to a decomposition approach. The study offers practical guidance for selecting the solution method to use based on instance characteristics and the scalability of the method.

4 - Novel Stratified Algorithms for Sustainable Decision-Making Under Uncertainty

Shayan Sharifi, Wayne State University, Detroit, MI, United States, Melike Yildirim

In the rapidly evolving landscape of public healthcare, decision-makers face complex challenges due to conflicting objectives and dynamic weights for each objective. Traditional decision-making methods often fall short in such dynamic and uncertain environments. This study introduces two novel algorithms—Multi-Objective Stratified Decision-Making (MOSDM) and Sustainable Stratified Decision-Making (SSDM)—developed based on the Concept of Stratification (CST), aimed at addressing these challenges in public healthcare. The MOSDM algorithm significantly reduces computational complexity, making it suitable for large-scale applications and overcoming the limitations of existing stratification methods. Building on MOSDM, the SSDM algorithm incorporates sustainability principles, allowing decision-makers to consider future scenarios and fluctuating criteria—an essential feature in healthcare, where long-term sustainability is critical. These algorithms were validated through real-world case studies, where their performance was benchmarked against traditional decision-making approaches. The results demonstrate the algorithms' superior ability to integrate multiple objectives, adapt to changing environments, and promote sustainability. Key findings highlight their scalability, efficiency, and adaptability, showcasing their potential to enhance decision-making in healthcare policy, resource allocation, and long-term strategic planning.

5 - Sequentially Eliciting Distributional Information from a Human for Stochastic Integer Programs

John Nichols, Rensselaer Polytechnic Institute, Troy, NY, United States, Jennifer Pazour, Sandipan Mishra, John Mitchell

Stochastic programming problems have uncertain parameters, which can be represented as probability distributions. Expert elicitation protocols can be used to reveal additional information about these uncertain parameters. In this work, we develop a method to guide expert elicitations used to update stochastic input parameters. The method is designed to balance the need to create high-quality solutions without overburdening the human expert. This method uses statistical distances measures in a one-dimensional line search to determine how much an uncertain parameter's probability distribution can be perturbed before the objective function value changes. This is the stability region. The method recommends querying the human about the stochastic parameter with the smallest stability region. This is a general, numerical approach for stochastic integer programs. This method is compared to benchmark methods, including eliciting a human expert for information about all uncertain parameters.

6 - Defining and Solving Stochastic Programs with Equilibrium Recourse

Beau Groom, University of Tennessee, Knoxville, TN, United States, Mingzhou Jin

Applications of stochastic programming that involve modeling human behavior generally assume that independent agents choose to act in ways that are beneficial on the societal level. However, game theory shows that people follow their incentives to reach Nash equilibria at suboptimal societal outcomes. These interactions of self-interested human actors can cause large degradations in system performance that are not accounted for in traditional stochastic programming. To better model these

interactions and their impact on optimal first-stage decisions, we introduce stochastic programs with equilibrium-constrained recourse problems, along with a portfolio of exact and heuristic solution techniques. These algorithms are demonstrated and validated for an infrastructure investment planning problem with user equilibrium traffic assignment recourse.

7 - Managing Distributional Ambiguity in Stochastic Optimization Through a Statistical Upper Bound Framework

Shixin Liu, University of Michigan - Dearborn, Dearborn, MI, United States, Jian Hu

Stochastic optimization is often hampered by distributional ambiguity, where critical probability distributions are poorly characterized or unknown. Addressing this challenge, we introduce a new framework that targets the minimization of a statistical upper bound for the expected value of uncertain objectives, facilitating more statistically robust decision-making. Central to our approach is the Average Percentile Upper Bound (APUB), a novel construct that simultaneously delivers a statistically rigorous upper bound for the population mean and a meaningful risk metric for the sample mean. The integration of APUB into stochastic optimization not only fortifies the process against distributional ambiguity but also reinforces key data-driven decision-making attributes, such as reliability, consistency, and comprehensibility. Notably, APUB-enriched optimization problems feature tractability, with particular advantages in two-stage stochastic optimization with random recourse. Empirical demonstrations on two-stage product mix and multi-product newsvendor benchmark problems reveal the benefit of the APUB optimization framework, in comparison with conventional techniques such as sample average approximation and distributionally robust optimization.

8 - Learning Robust Decisions: Data-Driven Distributionally Robust Optimization with Neural Decision Rules

Zhongling Xu, The University of Texas at Austin, Austin, TX, United States

Traditional optimization often assumes perfect knowledge of underlying data distributions, a luxury rarely afforded in real-world applications. Data-driven distributionally robust optimization (DRO) offers a powerful framework to address this uncertainty by optimizing against the worst-case distribution within an ambiguity set, defined by available data. This talk will explore a novel approach to solving data-driven DRO problems by leveraging the expressive power of neural networks to represent optimal decision rules. Specifically, we will discuss how to parameterize the decision rule as a neural network, allowing us to learn complex, nonlinear, and robust policies directly from data. This methodology not only enhances the flexibility and applicability of DRO to high-dimensional and intricate problems but also provides a computationally tractable pathway to derive robust solutions. We will delve into the theoretical underpinnings and practical benefits of this neural network-based approach, demonstrating its effectiveness in generating decisions that are resilient to distributional shifts and model misspecification.

WB53

Building B Level 3 B313

Recent Topics in Learning and Decision Making

Invited Session

Revenue Management

Chair: Jiashuo Jiang, HKUST, Hong Kong, Hong Kong

1 - Taming the Curse of Dimension in Multi-Stage Decision-Making

Yifan Hu, Rutgers Statistics, New Brunswick, NJ, United States

We revisit the empirical approximation of multi-stage stochastic programming studied in Shapiro, 2006. For 3-stage problems, the previous results obtain an $O(\epsilon^{-4})$ complexity bound to obtain an ϵ approximation. It was thus inferred that the sample complexity of T-stage problem admits an exponential dependence on T, i.e., $O(\epsilon^{-2T+2})$. Such a conjecture forms a long-standing belief that multi-stage stochastic programming suffers from the curse of dimensionality. In this talk, we discuss a novel method for 3-stage problem that achieves an $O(\epsilon^{-2})$ complexity bound, indicating that the complexity bounds for T-stage problem can be improved. We further construct a novel empirical approximation of T-stage problem and establish a polynomial complexity bound in terms of $O(1/\epsilon)$.

2 - Online Order Fulfillment with Replenishment

Zi (Elaine) Ling, Booth School of Business, University of Chicago, Chicago, IL, United States, Jiashuo Jiang, Linwei Xin

In modern e-commerce and service operations, firms must jointly manage *inventory replenishment policies* and *real-time fulfillment algorithms* to maximize revenue under uncertainty. While each component has been studied extensively in isolation, their interaction remains underexplored. This paper investigates a fundamental operational question: **Which lever has a greater impact on system performance—replenishment or fulfillment?**

We model the system as a one-location fulfillment problem (equivalently, a multi-secretary problem) with lost sales and stochastic customer arrivals, each offering heterogeneous rewards. Replenishment follows either a base-stock or constant-order policy, while real-time fulfillment decisions are made using online algorithms. Our core performance metric is the long-run average revenue per replenishment cycle, evaluated across all combinations of these policies and algorithms.

We show that, when the replenishment cycle length is large, the *total regret of online fulfillment* remains in the same order as in a single-cycle problem—even in the presence of replenishment—revealing a form of *regret stability*, which also extends to the multi-location setting. We further develop a regret-based framework for quantitatively comparing fulfillment and replenishment decisions, characterizing when optimizing the replenishment policy has greater impact on revenue than improving the fulfillment algorithm alone. Motivated by examples where online myopic algorithms underperform, we introduce a novel *look-ahead online algorithm* that anticipates future replenishment and demand. Numerical experiments verify that this algorithm outperforms myopic baselines.

Our results offer both theoretical and managerial insights into situations where inventory replenishment policies are more influential and vice versa.

3 - Beyond Non-Degeneracy: Revisiting Certainty Equivalent Heuristic for Online Resource Allocation

Wenjia Wang, HKUST (GZ), Guangzhou, China, People's Republic of, Yilun Chen

The Certainty Equivalent (CE) heuristic is a widely-used algorithm for various online resource allocation problems in OR and OM. Despite its popularity, existing theoretical guarantees of CE are limited to settings satisfying restrictive fluid regularity conditions, particularly, the non-degeneracy conditions, under the widely held belief that the violation of such conditions leads to performance deterioration and necessitates algorithmic innovation beyond CE.

In this work, we conduct a refined performance analysis of CE within the general framework of online linear programming. We show that CE achieves uniformly near-optimal regret (up to a polylogarithmic factor in T) under only mild assumptions on the underlying distribution, without relying on any fluid regularity conditions. Our result implies that, contrary to prior belief, CE effectively beats the curse of degeneracy for a wide range of problem instances with continuous conditional reward distributions, highlighting the distinction of the problem's structure between discrete and non-discrete settings. Our explicit regret bound interpolates between the mild $(\log T)^2$ regime and the worst-case $\sqrt{T}$ regime with a parameter β quantifying the minimal rate of probability accumulation of the conditional reward distributions, generalizing prior findings in the multisecretary setting.

To achieve these results, we develop novel algorithmic analytical techniques. Drawing tools from the empirical process theory, we establish strong concentration analysis of the solutions to random linear programs, leading to improved regret analysis under significantly relaxed assumptions. These techniques may find potential applications in broader online and offline stochastic decision-making contexts.

4 - Empowering LLM Alignment with Choice Modeling

Yuxuan Tang, National University of Singapore, Singapore, Singapore, Yifan Feng

LLM alignment is the process of making LLMs behave in ways that are consistent with human intentions and values. It's crucial because unaligned models may produce harmful, biased, or undesired outputs. Human preference data is central to alignment, helping steer models toward more desirable responses. While traditional approaches

like RLHF and DPO rely on pairwise preference data, our work explores a more expressive format—ranked choice data—which allows for more efficient and informative training.

WB55

Building B Level 3 B314

Consumer Choice Behavior

Invited Session

Revenue Management

Chair: Gustavo Vulcano, Universidad Torcuato Di Tella, Buenos Aires, 1426, Argentina

1 - Adaptive Two-Sided Assortment Optimization: Revenue Maximization

Mohammadreza Ahmadnejadsaein, Cornell Tech, Cornell University, New York, NY, United States, Omar El Housni

We study adaptive two-sided assortment optimization for revenue maximization in choice-based matching platforms. The platform has two sides of agents, an initiating side, and a responding side. The decision-maker sequentially selects agents from the initiating side, shows each an assortment of agents from the responding side, and observes their choices. After processing all initiating agents, the responding agents are shown assortments and make their selections. A match occurs when two agents mutually select each other, generating pair-dependent revenue. Choices follow Multinomial Logit (MNL) models. This setting generalizes prior work focused on maximizing the number of matches under submodular demand assumptions, which do not hold in our revenue-maximization context.

Our main contribution is the design of polynomial-time approximation algorithms with constant-factor guarantees. In particular, for general pairwise revenues, we develop a randomized algorithm that achieves a $(\frac{1}{2} - \epsilon)$ -approximation in expectation for any $\epsilon > 0$. The algorithm is static and provides guarantees under various agent arrival settings, including fixed order, simultaneous processing, and adaptive selection. When revenues are uniform across all pairs involving any given responding-side agent, the guarantee improves to $(1 - \frac{1}{e} - \epsilon)$. In structural settings where responding-side agents share a common revenue-based ranking, we design a simpler adaptive deterministic algorithm achieving a $\frac{1}{2}$ -approximation. Our approach leverages novel linear programming relaxations, correlation gap arguments, and structural properties of the revenue functions.

2 - Contextual Markov Chain Choice Model

Kai Tan, Stanford University, Stanford, CA, United States, Elynn Chen, Gustavo Vulcano, Jiawei Zhang

In this work, we propose a contextual Markov Chain Choice Model (MCCM) that captures heterogeneity in customer purchasing behavior by incorporating both customer and product features. We develop an Expectation-Maximization (EM) algorithm to perform maximum likelihood estimation and establish the convergence of the log-likelihood function. In addition, we introduce a linear programming approach to compute revenue-maximizing assortments, enabling personalized assortment decisions for individual customers. Extensive simulation studies show that our contextual MCCM is noticeably competitive with existing benchmark methods in both modeling customer choices and revenue optimization.

3 - On the Estimation of a Markov Choice Model from Partial Inventory Data

Gustavo Vulcano, Universidad Torcuato Di Tella, Buenos Aires, Argentina, Charles Thraves, Ivan Vidal

A seller observes the inventory levels of substitute products at the beginning and end of each day over multiple days. Each day, customers arrive sequentially and either purchase one product or opt for the no-purchase alternative. Each customer has a probability distribution over first-choice preferences and, if their preferred product is unavailable, transitions to another option according to a Markov chain discrete choice model. While the total number of arriving customers is known, neither the specific product purchased by each customer nor the choice set they faced is observed. This work extends previous studies by relying on less detailed information about customer behavior and observed outcomes. We use the Expectation-Maximization (EM) algorithm to estimate the model parameters, using Metropolis-Hastings (MH) methods in the E-step to sample purchase paths consistent with inventory observations. The M-step admits closed-form solutions. We solve numerical instances to assess the accuracy and performance of the parameter estimates. Our results show that, although the true parameters are not always identifiable, the log-likelihood achieved with the estimated parameters is very close to that obtained using the true parameters. This finding also holds when evaluating simulated data in terms of final inventories, stock-out probabilities, and fill rates.

4 - Blind and near-Optimal Dynamic Assortment Optimization with Reusable Resources

Zerui Wu, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Ran LIU, Xu Sun

We study dynamic assortment optimization with reusable resources in blind settings, where both customer arrival rates and resource rental durations are unknown. This challenge arises in environments such as cloud computing, healthcare, and hospitality, where resources are rented for random durations and demand patterns may fluctuate. We propose a simple, blind, state-dependent policy that selects assortments based solely on the current system state and known resource attributes, such as expected revenue and customer choice probabilities. The policy requires no estimation of arrival or rental return rates and adapts naturally to slowly time-varying conditions. It achieves an absolute revenue loss of $O(\sqrt{n})$ in general systems and a constant $O(1)$ loss in overloaded regimes, where n is the total number of resources, matching the best-known performance bounds in the literature under comparable assumptions. Theoretical guarantees are derived using an exponential potential function that induces self-correcting behavior without explicit learning. Numerical experiments support the theoretical bounds and demonstrate robust performance across a range of operating regimes.

WB56

Building B Level 3 B315

Learning in Games and Strategic Environments

Invited Session

Auctions and Market Design

Chair: Kunhe Yang, UC Berkeley, Berkeley, CA, United States

1 - Calibrated Stackelberg Games: Learning Optimal Commitments Against Calibrated Agents

Chara Podimata, MIT, Newton, MA, United States, Nika Haghtalab, Kunhe Yang

In this paper, we introduce a generalization of the standard Stackelberg Games (SGs) framework: Calibrated Stackelberg Games (CSGs). In CSGs, a principal repeatedly interacts with an agent who (contrary to standard SGs) does not have direct access to the principal's action but instead best-responds to calibrated forecasts about it. CSG is a powerful modeling tool that goes beyond assuming that agents use ad hoc and highly specified algorithms for interacting in strategic settings and thus more robustly addresses real-life applications that SGs were originally intended to capture. Along with CSGs, we also introduce a stronger notion of calibration, termed adaptive calibration, that provides fine-grained any-time calibration guarantees against adversarial sequences. We give a general approach for obtaining adaptive calibration algorithms and specialize them for finite CSGs. In our main technical result, we show that in CSGs, the principal can achieve utility that converges to the optimum Stackelberg value of the game both in finite and continuous settings, and that no higher utility is achievable. Two prominent and immediate applications of our results are the settings of learning in Stackelberg Security Games and strategic classification, both against calibrated agents.

2 - Learning Local Stackelberg Equilibria from Repeated Interactions with a Learning Agent

Nivasini Ananthkrishnan, University of California, Berkeley, Berkeley, CA, United States, Yuval Dagan, Kunhe Yang

Motivated by the question of how a principal can maximize its utility in repeated interactions with a learning agent, we study repeated games between a principal and an agent employing fictitious play. Prior work has shown that computing or even approximating the global Stackelberg value in similar settings can require an exponential number of rounds in the size of the agent's action space, making it computationally intractable. In contrast, we shift focus to the computation of local Stackelberg equilibria and introduce an algorithm that, within the smoothed analysis framework, constitutes a Polynomial Time Approximation Scheme (PTAS) for finding an ϵ -approximate local Stackelberg equilibrium. Notably, the algorithm's runtime is polynomial in the size of the agent's action space yet exponential in $(1/\epsilon)$ —a dependency we prove to be unavoidable.

3 - Nearly-Optimal Bandit Learning in Stackelberg Games with Side Information

Keegan Harris, Carnegie Mellon University, School of Computer Science, Pittsburgh, PA, United States

We study the problem of online learning in Stackelberg games with side information between a leader and a sequence of followers. In every round the leader observes contextual information and commits to a mixed strategy, after which the follower best-responds. We provide learning algorithms for the leader which achieve $O(T^{1/2})$ regret under bandit feedback, an improvement from the previously best-known rates of $O(T^{2/3})$. Our algorithms rely on a reduction to linear contextual bandits in the utility space: In each round, a linear contextual bandit algorithm recommends a utility vector, which our algorithm inverts to determine the leader's mixed strategy. We extend our algorithms to the setting in which the leader's utility function is unknown, and also apply it to the problems of bidding in second-price auctions with side information and online Bayesian persuasion with public and private states. Finally, we observe that our algorithms empirically outperform previous results on numerical simulations.

4 - Full Swap Regret and Discretized Calibration

Princewill Okoroafor, Cornell University, Ithaca, NY, United States, Maxwell Fishelson, Robert Kleinberg, Renato Paes Leme, Jon Schneider, Yifeng Teng

We study the problem of minimizing swap regret in structured normal-form games. Players have a very large (potentially infinite) number of pure actions, but each action has an embedding into d -dimensional space and payoffs are given by bilinear functions of these embeddings. We provide an efficient learning algorithm for this setting that incurs at most $O(T^{(d+1)/(d+3)})$ swap regret after T rounds.

To achieve this, we introduce a new online learning problem we call *full swap regret minimization*. In this problem, a learner repeatedly takes a (randomized) action in a bounded convex d -dimensional action set $\mathcal{K}$ and then receives a loss from the adversary, with the goal of minimizing their regret with respect to the *worst-case* swap function mapping $\mathcal{K}$ to $\mathcal{K}$. For varied assumptions about the convexity and smoothness of the loss functions, we design algorithms with full swap regret bounds ranging from $O(T^{d/(d+2)})$ to $O(T^{(d+1)/(d+2)})$.

Finally, we apply these tools to the problem of online forecasting to minimize calibration error, showing that several notions of calibration can be viewed as specific instances of full swap regret. In particular, we design efficient algorithms for online forecasting that guarantee at most $O(T^{1/3})$ ℓ_2 -calibration error and $O(\sqrt{\epsilon} T, T^{1/3})$ *discretized-calibration* error (when the forecaster is restricted to predicting multiples of ϵ).

WB58

Building B Level 3 B317

Healthcare Operations I

Contributed Session

Contributed

Chair: Esmail Bahalkeh, University of New Hampshire, Durham, NH, United States

1 - A Patient-Centered Hospital Procurement of Physician Preferred Implants

Renato de Matta, University of Iowa, Iowa City, IA, United States

This paper studies the hospital's procurement of physician preferred implants where both hospital surplus (profit) and patient surplus (value) are maximized. For better flexibility with minimal revision risk to patients, we introduce the substitution of another device that belongs in the category of the physician-preferred implant. This allows for improving the hospital surplus through larger revenues and volume discounts, although it implies that there is an additional cost to train the physician who may need to use the substituted implant and associated surgical procedure for safety and quality assurance. We model the procurement problem as a single period multiproduct stochastic programming problem with substitution. The problem is NP-hard. We propose two heuristic procedures for finding good feasible solutions and for estimating the optimality bound. The results of our experiments using patient-reported costs and total hip replacement outcomes from the medical literature show that a modest increase in flexibility can improve the hospital's financial performance, produce better patient outcomes, and reduce the negative effects of demand uncertainty.

2 - Streamlining Hospital Operations with Interactive Optimization and AI-Powered BI Tools

Deepti Bahel, VigourSoft Global Solutions, San Jose, CA, United States, Vinay Manglani

This presentation showcases an interactive hospital business intelligence (BI) application developed using Streamlit, designed to support real-time, data-driven decision-making in healthcare operations. The tool enables stakeholders to explore and analyze hospital data across key areas such as billing, patient length of stay, and anomaly detection—without requiring programming expertise.

At its core, the application offers optimization capabilities using linear programming (LP), integer programming (IP), and greedy heuristics to help hospital administrators identify cost-efficient patient management strategies. Users can apply constraints like long-stay thresholds, weekend admission limits, and anomaly risk to simulate and compare different approaches.

The platform integrates machine learning models—such as XGBoost and Random Forest—to classify patient stays and predict outcomes, while Isolation Forest algorithms help detect potential billing anomalies. Model explainability is enhanced with SHAP values, making it easier to understand the drivers behind predictions.

What makes this tool particularly accessible is its conversational AI interface. Users can query the data in plain English, and the system returns executable Python code, visualizations, and narrative summaries generated by GPT. With features like CSV upload, filters, and result exports, the app provides a seamless user experience.

This talk will walk through the design, key features, and practical use cases of the tool, demonstrating how operations research, machine learning, and explainable AI can work together to improve transparency and efficiency in hospital management.

3 - The Impact of DIP Policy on Patient Hospitalization Costs and Hospital Admission Practices

Nuotong Yang, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Hui Lin, Guohua Wan

Based on inpatient data from DIP pilot regions, this study employs a difference-in-differences approach to empirically examine the policy's impact on medical costs and their structure, while analyzing adjustments in hospital admission practices. The findings reveal that the DIP policy significantly reduces hospitalization costs in secondary hospitals and decreases examination and testing expenses for inpatients in tertiary hospitals, demonstrating notable heterogeneity across hospital tiers. The study suggests that the DIP policy can leverage economic incentives to motivate healthcare institutions to control costs endogenously, thereby curbing unreasonable growth in medical expenditures.

4 - Designing Specialist-Response Policies in Hospital Emergency Departments

Emily Fainman, Texas State University, San Marcos, TX, United States, Beste Kucukyazici, Zhankun Sun

Emergency Department (ED) overcrowding is a global healthcare challenge, significantly exacerbated by delays in specialist consultations (SCs). These delays, measured as Request-to-Realization (R2R) times—the interval from SC request to specialist arrival—often exceed targets and prolong ED length of stay (LoS), thereby compromising patient outcomes and operational efficiency. This study addresses the design of specialist-response policies to mitigate R2R delays without increasing staffing burdens. Leveraging data from over one million ED visits across eleven hospitals in Montreal, we employ empirical analysis, stochastic modeling, and simulation to assess the impact of SC delays and develop optimal specialist scheduling strategies.

We model SC processes as $Mt/G(T)/1$ queues to capture time-varying arrival patterns and derive an analytical, closed-form solution for optimal fixed-time specialist schedules. These are evaluated against conventional on-call and on-site policies to identify the most efficient policy mix for each hospital setting. Furthermore, we integrate specialist availability into triage prioritization by developing a predictive model to assess SC needs at triage, aligning patient assessment with scheduling constraints.

Simulation studies in both tertiary and secondary hospitals validate the approach, demonstrating reductions in R2R times by 25–42% and ED LoS by over 10%, without increasing the frequency of specialist visits. This research introduces a novel, time-dependent queuing framework for healthcare operations, offering actionable decision-support tools for hospital administrators. Our findings have broad applicability in managing shared or rotating medical staff, thereby enhancing ED responsiveness and patient care across diverse hospital environments.

5 - Nurse Staffing and Hospital Outcomes: a Strategic Investment in Quality and Sustainability

Esmail Bahalkeh, University of New Hampshire, Durham, NH, United States, Dinesh Pai

Rising healthcare costs and the ongoing shift toward value-based care necessitate a thoughtful and critical reassessment of how hospitals allocate their resources—particularly in relation to nursing staff, who remain the backbone of high-quality patient care. This longitudinal study (2000–2023) analyzes data from acute care hospitals across Pennsylvania to explore the complex, multidimensional relationships between nurse staffing levels (RN ratio), skill mix (proportion of RN), financial performance, quality of care, and length of stay (LOS). Using robust statistical methods, we examine how variations in staffing influence key clinical and economic outcomes over time. Our findings reveal a nuanced and often paradoxical landscape. Hospitals with a higher RN ratio experience significantly better quality outcomes. However, these clinical benefits are accompanied by increased operational costs and diminished operating margins, highlighting a financial trade-off that administrators must carefully consider. Meanwhile, a higher skill mix is associated with favorable impacts on readmissions and length of stay but does not show a statistically significant relationship with financial performance, suggesting that improvements in workforce composition may enhance care efficiency without clear cost consequences. By illuminating both the clinical advantages and financial implications of investing in a skilled nursing workforce, this research offers timely, data-driven insights for healthcare leaders and policymakers. It emphasizes the importance of strategic, evidence-based staffing decisions to achieve both financial sustainability and optimal patient outcomes within an increasingly demanding and value-conscious healthcare environment.

WB60

Building B Level 3 B319

Advances in Deep Learning

Contributed Session

Contributed

Chair: Marat Ibragimov, Emory University, Atlanta, GA, United States

1 - Explaining Black Box Models: a Framework Comparison on Loan Data Using Prosper Marketplace Loan Data

Rongbin Ye, Synchrony Bank(Self Representing), Chicago, IL, United States, Jiaqi Chen

Assessing credit risk in finance is challenging due to the need to balance powerful machine learning (ML) predictions with regulatory demands for transparency and fairness (e.g., OCC, CFPB). While simple models are transparent but less accurate, complex "black-box" models lack explainability. This project tackles this by developing and comparing Logistic Regression, Random Forest, and Neural Network models for credit risk on the Prosper Marketplace dataset, enhanced with macroeconomic data. We used LIME to explain individual predictions, especially for the Neural Network. Results show complex models predict better, particularly for defaults, and LIME effectively reveals their decision processes. We also propose a new five-part framework (Inherent Interpretability, Global Explanations, Local Explanations, Consistency, Complexity) to evaluate model explainability beyond just accuracy. This research demonstrates that high-performance ML models can be used in regulated finance by employing modern explainability methods and provides a structured way to weigh performance against interpretability.

3 - Data Driven Delivery Delay Prediction Using a Deep Learning Framework

Stefan Faulkner, Georgia Institute of Technology, Atlanta, GA, United States, Reza Zandehshahvar, Vahid Eghbal Akhlaghi, Pascal Van Hentenryck

Accurate and timely predictions of delivery times play a crucial role in enhancing supply chain operational efficiency and ensuring customer satisfaction. Yet, numerous challenges exist in making precise predictions, including complex logistics data, geographical differences, and imbalanced delivery status distributions. This paper introduces a two-phase deep learning approach for predicting delivery delays using a classification-then-regression setting. The models are equipped with uncertainty quantification through conformal prediction to enhance the trustworthiness and enable informed-decision making.

The approach begins by analyzing historical delivery data to develop a classification system that categorizes shipments into three forms of groups: on-time, delayed, and those with uncertain status. For shipments identified as potentially delayed or uncertain, a second-phase regression model calculates the expected delay time. Both stages include conformal prediction to ensure well-calibrated predictions with reliable confidence levels. The framework accounts for geographical variations by training location-specific models based on the shipment sources.

The proposed method employs deep learning with specialized embedding layers for numerical and categorical variables and is compared against conventional machine learning approaches. Testing on an extensive real-world dataset containing over 10 million shipment records from an industry logistics provider demonstrates that multi-layer perceptron model with linear embeddings surpasses alternative approaches, yielding F1-scores ranging from 0.35 to 0.49 across various locations despite significant class imbalances. The regression component also performs effectively, achieving mean absolute errors under 1 day for delayed shipments and under 1.5 days across all shipments.

4 - An Xr-Enabled Deep Learning Framework for Driver Distraction Detection and Monitoring

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

According to the World Health Organization (WHO), approximately 1.19 million people die annually in road traffic accidents, with an additional 20 to 50 million sustaining non-fatal injuries, often leading to long-term disabilities. A significant portion of these accidents is attributed to driver distraction, particularly from mobile phone use and drowsiness—two of the most pressing safety concerns in modern transportation. This study addresses these challenges by utilizing the publicly available *Distracted Driver Dataset V2*, which includes side-view RGB images of 44 drivers performing 10 distinct actions related to distracted driving. To enhance the dataset's coverage and relevance, we introduce an additional class representing drowsy drivers, acknowledging that drowsiness was involved in approximately 17.6% of fatal crashes between 2017 and 2021. We implement transfer learning using a lightweight Deep Convolutional Neural Network (DCNN) based on the MobileNet architecture. The proposed model achieves a classification accuracy of 93.60% and an F1-score of 0.94 across all classes. Beyond training and evaluation, we deploy the trained model on an Extended Reality (XR) headset to enable real-time driver monitoring from a passenger-side view. The XR integration offers enhanced situational awareness and posture visibility, improving detection accuracy for subtle or partially occluded distractions. By fusing lightweight AI models with immersive sensing platforms, our

approach offers a practical and scalable solution for real-time distracted driver classification. The results highlight the potential of combining transfer learning with XR technologies for advanced, adaptive, and context-aware driver safety systems.

5 - A Human-on-the-Loop LLM Framework for Adaptive Quality Prediction Systems for Industry 5.0

Minsik Kim, Kyung Hee University, Yongin, Korea, Republic of, Jae-Yoon Jung

To realize autonomous manufacturing, it is essential to develop technologies that minimize human intervention across the production process. This study proposes a semi-autonomous and adaptive quality prediction system using large language models (LLMs). The proposed framework aims to address two key challenges in intelligent manufacturing environments: real-time adaptability to data distribution shifts and model interpretability.

This study designs an LLM-based pipeline that learns from production data and extracts rule-based explanations for quality prediction. The system offers clear reasoning behind predictions through explainable AI, while its automated quality monitoring mechanism continuously observes model performance. Upon detecting data shifts, the model is dynamically updated to maintain long-term prediction accuracy.

This approach demonstrates that LLMs can be effectively applied not only to natural language tasks but also to mission-critical domains such as manufacturing. Experimental results using real-world production datasets show that the proposed system maintains robust predictive performance under changing conditions, while also providing human-understandable explanations.

These findings highlight the potential of integrating industrial AI with LLMs to develop interpretable and adaptive autonomous quality control systems. Ultimately, the proposed method supports greater automation and resilience in manufacturing by reducing dependency on manual oversight.

6 - Improving Targeting Policies Using Transfer Learning

Marat Ibragimov, Emory University, Atlanta, GA, United States, Duncan Simester, Artem Timoshenko

Firms can enhance the profitability of marketing campaigns by targeting different customers with different actions based on the incremental effects. However, firms typically estimate the incremental effects using training data that closely resembles the focal campaign, and this data is often scarce. We propose a method to augment focal training data with information from different marketing campaigns, even when the campaigns differ in content, audience, and timing. Our approach accounts for campaign and customer differences and explicitly incorporates uncertainty in the measurement of treatment effects using a Bayesian probabilistic matrix factorization framework. Scalability and computational efficiency are achieved through a combination of closed-form conditional distributions and gradient-based MCMC sampling. We validate the method using simulated data and a large-scale field experiment, and demonstrate substantial improvements in predictive accuracy and campaign profitability. By leveraging information across campaigns, the proposed approach amplifies the value of firms' marketing data and offers a scalable and robust solution for designing targeting policies.

WB63

Building B Level 4 B403

AI in Financial and Technical Applications

Contributed Session

Contributed

Chair: Meijing Zhang, Singapore University of Technology and Design, Singapore, Singapore

1 - Treemark++: Robust and High-Capacity Image Watermarking for Diffusion Models

Yaopeng Wang, Southeast University, Nanjing, China, People's Republic of

As diffusion-based generative models are increasingly applied in content creation, image watermarking has become a critical tool for ensuring traceability and copyright protection. However, existing methods struggle to balance robustness, imperceptibility, and information capacity. In this work, we propose TreeMark++, an enhanced watermarking framework built upon Tree-Ring Watermarking. TreeMark++ introduces a single-layer ring structure with multi-bit embedding, significantly improving capacity and matching efficiency in multi-user scenarios. To address synchronization issues under geometric transformations, we incorporate a header-aware design and redundancy encoding, enabling accurate information extraction even under distortions such as JPEG compression, cropping, and large-angle rotation. Experimental results demonstrate that TreeMark++ achieves high watermark detection and extraction accuracy while preserving semantic fidelity across diverse datasets.

2 - Smart Home Service System Development Framework Based on Generative AI

JAESEOK YUN, Korea University, Sejong, Korea, Republic of, Sung Yeon Kim, Seonghyeon Han, Jinmin Kim

This study proposes a development framework for smart home service systems based on generative artificial intelligence (Generative AI) and analyzes its effects. Traditional smart home systems have faced limitations in providing personalized services due to a lack of interoperability among devices, command-based interaction, and non-standardized data management. Focusing on the rapid advancement of Generative AI technologies, this research presents a framework that enables more intelligent and integrated smart home services. The proposed framework incorporates data standardization, user requirement-based system design, and personalized service strategies, while also aligning with global smart home standards such as Matter. Additionally, the study empirically examines the impact of Generative AI-based smart home systems on users' intention to use the services and their residential satisfaction. The analysis reveals perceptual structures based on differences in product-service quality across smart home components, suggesting the potential for a new classification framework from the perspective of Product-Service Systems (PSS). Notably, the proposed framework also supports the advancement of smart home testing and certification systems, contributing to enhanced interoperability and system reliability. This study not only highlights the applicability of Generative AI in the smart home domain but also provides technological and policy implications for implementing sustainable smart homes, improving user experience, and enabling data-driven decision-making.

3 - Comparing Deep Learning Models and LLMs for Airport Sentiment Analysis on Twitter Data

Dirk Van den Poel, Ghent University, Ghent, Belgium, Belgium, Maxim Devriendt, Valentin Janclaes

This study presents a structured and scalable pipeline for extracting actionable insights on airport service experiences from unstructured Twitter data. We use a staged setup to filter over two million tweets into a curated subset of relevant, airport-specific content, enabling fine-grained sentiment evaluation per airport service category.

We benchmark both deep learning models (e.g., ModernBERT, BiLSTM) and a range of large language models (LLMs), including lightweight (e.g., Mistral 7B, Qwen 2.5 3B) and heavyweight variants (e.g., LLaMa 3.3 70B, DeepSeek 67B). In early pipeline tasks, such as relevance filtering and entity classification, ModernBERT (an encoder-based model) outperforms LLMs in both F1-score and computational efficiency. However, in the airport service topic modeling stage, where multi-label complexity and semantic nuance are essential, LLMs, particularly heavyweight ones, exceed deep-learning baselines. The final stage performs sentiment analysis across the airport service topics.

Our approach differs from prior work in three ways: (1) it enables multi-label classification across multiple airport service topics per tweet, rather than restricting to a single label; (2) it introduces a modular, step-by-step pipeline tailored to noisy Twitter data, transferable to other social media sources; and (3) it provides comprehensive benchmarking of LLMs versus deep learning throughout the pipeline.

We demonstrate that social media data, when processed via a structured pipeline, may yield rich, near real-time insights into passenger experience, offering a valuable complement to traditional airport service survey methods and enabling more agile, data-driven airport management.

4 - Generative AI-Based Optimization Agent for Nonlinear Programming

Saif Kazi, Los Alamos National Laboratory, Santa Fe, NM, United States

Nonlinear Constrained Optimization a.k.a Nonlinear Programming (NLP) is an important tool in solving engineering problems [Optimal Control, Optimal Power Flow, Nonlinear Model Predictive Control etc.] with applications in multiple domains such as operations research, finance etc. Deterministic methods to solve NLPs are generally based on using the optimality criterion known as Karush-Kuhn-Tucker (KKT) conditions. Algorithms such as Interior-point methods (IPM) use Newton-Raphson to solve the nonlinear KKT conditions. This particular approach suffers from convergence issues due to the local convergence property of Newton's algorithm making it extremely important to provide good initial guess to the optimization algorithm for fast convergence.

In this study, we propose building an optimization agent based on mix of LLM and Neural Network architecture which first transforms NLP problems written in text format to the KKT conditions. In second step, the Nonlinear KKT Equation system is solved using the Neural network architecture and an adaptive learning rate. This step doesn't require training of neural network using artificial dataset and also overcomes the local convergence issue associated with the classical Newton-Raphson method. We test our approach on benchmark NLP problems to showcase and compare it's performance with conventional NLP solvers.

5 - Multi-Agent Language Agent Tree Search for Adaptive Financial Decision-Making

Sorouralsadat Fatemi, University of Illinois at Chicago, Chicago, IL, United States

In this work, we introduce **Language Agent Tree Search (LATS)**, a novel multi-agent framework that integrates the reasoning power of large language models (LLMs) with structured decision-making for adaptive financial tasks. LATS extends Monte Carlo Tree Search (MCTS) by embedding language agents at decision nodes, allowing each agent to generate, reflect, and critique financial actions—such as buy/hold/sell strategies—based on textual, tabular, and temporal data. Unlike prior agent-based approaches that rely on static heuristics or black-box models, our architecture coordinates multiple specialized agents, including reflection and forecasting agents, to dynamically explore and prune the decision tree through iterative, language-guided rollouts

6 - Transmode-LLM: Feature-Informed Natural Language Modeling with Domain-Enhanced Prompting for Travel Behavior Modeling

Meijing Zhang, Singapore University of Technology and Design, Singapore, Singapore, Ying Xu

Understanding traveler behavior and accurately predicting travel mode choice are at the heart of transportation planning and policy making. Traditional methods relying on raw numbers and structured feature representations have limitations on capturing the complex interdependency and qualitative factors that may impact on travel behavior in the real-world, particularly the rich contextual nuances underlying individual decision-making processes. Large language models (LLMs) with promising capabilities for understanding contextual information across domains provides new pathways for travel behavior modeling. In this study, we propose, TransMode-LLM, an innovative framework designed to predict travel modes from natural language descriptions of travelers and their trips. We start by analyzing the importance of features to identify and select key impacting factors (i.e. individual, household and trip characteristics) to enrich context for decision-making. To enhance the performance of LLMs for transportation-specific tasks, we propose a domain-enhanced prompting strategy that incorporates standardize mode definitions. We further explore various learning paradigms (zero-shot and one/few-shot learning) to understand their impact on travel mode prediction using natural language. Finally, we build an evaluation system to compare the performance of the proposed LLM-based approach against state-of-the-art traditional models. Extensive experiments are conducted on the real-world travel survey dataset and the results demonstrate the competitive performance of LLM-based approach such as prediction accuracy compared to the traditional methods. This study advances the application of LLMs in travel behavior modeling, providing promising and valuable insights for both academic research and transportation policy-making in the future.

WB64

Building B Level 4 B404

Supply Chain Strategy

Contributed Session

Contributed

Chair: Zhili Yuan, Shanghai Jiao Tong University, Shanghai, N/A

1 - Balancing Cost and Carbon Emissions: Smart Scheduling with Non-Identical Parallel Machines and Mixed Energy Sources in Manufacturing

Sang Jin Kweon, Yonsei University, Seoul, Korea, Republic of

In this talk we address a production scheduling problem involving non-identical parallel machines in a high-mix, low-volume manufacturing setting. We focus on two key objectives: minimizing production-related costs and reducing carbon emissions, while also accounting for machines powered by various energy sources, each with different costs and emission profiles. To tackle this, we develop a bi-objective mixed-integer programming model that supports green manufacturing strategies by finding cost-effective, low-emission production schedules. The proposed model is tested using real data from a manufacturer in Ulsan, South Korea, demonstrating how flexible machine assignments and energy source choices can jointly optimize cost and environmental impact. Our analysis reveals key trade-offs, such as shifts in energy preference under varying carbon emission limits. We also conduct sensitivity analyses on energy prices, showing that coal tends to offer more cost stability than natural gas during price fluctuations. Ultimately, we identify optimal energy source combinations under different emission constraints, providing insights for decision-makers striving to balance economic and environmental goals in production planning.

2 - When and How Should South Korea Achieve Carbon Neutrality? Understanding Public Preferences and Generation Differences for Carbon Neutrality Policy

Sukhoon Lee, Kyung Hee University, Yongin, Korea, Republic of, Hyunhong Choi

This study quantitatively examined public preferences and generational differences related to South Korea's carbon neutrality policy. The survey focused on three key attributes, including the target year (which included levels such as 2045, 2050, 2060, or 2070), the strategy (such as front-loaded, evenly distributed, or back-loaded), and the annual carbon neutrality tax (ranging from about 7 to 50 USD). An online survey was conducted in May 2024 with 645 adults aged 20 to 69 nationwide, utilizing a choice experiment and a hierarchical Bayesian logit model. Results showed the public most preferred 2050 as the target year, followed by 2045, 2060, and 2070. Similarly, the back-loaded strategy was least preferred. Additionally, the relative importance revealed that the combined weight of the target year and strategy accounted for approximately 40%, demonstrating that the public places significant emphasis on the strategic combination of when and how carbon neutrality should be achieved. A distinct generational pattern emerged, showing that older adults (aged 30 to 69) were even less likely than younger adults (aged 20 to 29) to prefer postponing both the target year and the achievement strategy, which would shift the burden to future generations. These findings suggest that implementing an additional carbon neutrality tax should involve differential tax rates by generation to ensure intergenerational equity is reflected in policy design. By analyzing generational differences in preferences for carbon neutrality policies, this study enhances understanding of intergenerational disparities and provides a foundation for developing more valid and participatory carbon neutrality policies in South Korea.

3 - Mitigating Disruptions in Semiconductor Supply Chains: Stability Improvement vs. Backup Procurement

Zhili Yuan, Shanghai Jiao Tong University, Shanghai, China, People's Republic of, Xiaofeng Shao, Jian Wang

Semiconductor supply chain disruptions often lead to prolonged recovery. Manufacturers face strategic trade-offs between stabilization improvement investments and backup procurement activation to enhance supply resilience. This study employs a game-theoretic approach to examine the optimal decision-making strategies under emergency procurement constraints. Results show that when emergency procurement is not allowed, investment only is viable under low stabilization costs and the disruption risk below an upper threshold, while reservation only becomes preferable when costs are high and the risk exceeds a lower threshold. Our findings further reveal that investment can sometimes deter backup procurement initiation, with a combined strategy optimal only under high-risk conditions. Notably, when permissible, emergency procurement proves incompatible with reservation and is adopted only if the disruption probability is below a defined value. Additionally, while investment increases the preference for emergency procurement, it does not affect purchase volumes. We extend assumptions yielding consistent insights. Moreover, when the supply

stability level is endogenous, a combined investment and emergency procurement strategy may surprisingly be optimal under high-risk scenarios. In contexts with risk-averse manufacturers and product competition, the risk threshold for selecting backup supply based solely on reservation increases, whereas it remains unchanged when backup supply offers procurement flexibility.

WB65

Building B Level 4 B405

Artificial Intelligence for Decision-Making in Manufacturing Systems

Invited Session

Artificial Intelligence

Chair: Lingxiang Yun, University of North Carolina at Charlotte, Charlotte, NC, United States

Co-Chair: Chen LI, Nanjing University, Nanjing, N/A

1 - Physics-Centric Machine Learning to Improve Fatigue Performance Prediction of Laser Powder Bed Fusion

Jia Liu, Auburn University, Auburn, AL, United States

Understanding the fatigue performance and accurately predicting the fatigue life of laser powder bed fusion (L-PBF) parts remains a pressing challenge due to its complex failure mechanisms, time-consuming tests, and limited fatigue testing data. Meanwhile, machine learning provides a promising alternative approach to investigating such complex relationships and using surrogate models to enable sufficient prediction. With previous knowledge in both fracture mechanics and machine learning, we propose a physics-centric machine learning method to investigate the relationship between fatigue performance and surface quality and improve fatigue life prediction. The fatigue life of additively manufactured metallic parts in the unmachined surface condition is typically shorter than those in machined surfaces, primarily due to the presence of surface micro-notches. Specifically, we utilize a fracture mechanics-based approach to identify the key surface features, which include the geometry of surface micro-notches, such as width, depth, opening angle, and radius of curvature, from Ti-6Al-4V L-PBF coupons. Then, we leverage the structure of a linear mixed model and point process to create a novel, interpretable, and generalizable machine-learning model to model the complex relationships between fatigue performance and surface quality, aiming to achieve a better interpretation, smaller scatter, and more accurate prediction.

2 - Enhanced Counterfactual Explanations for Optimizing 3D Printing Parameters Using Shap and Nearest-Neighbor Constraints

Samar Saleh, Rutgers, The State University of New Jersey, Piscataway, NJ, United States, Weihong Guo, Yuebin Guo

Overheating in additive manufacturing (AM), particularly in Laser Powder Bed Fusion, impacts part quality and reliability. This study presents a novel framework for optimizing 3D printing parameters (power, speed, and printing sequence) using enhanced counterfactual explanations (CE).

The framework incorporates SHapley Additive exPlanations (SHAP) to identify key features influencing overheating, quantified by the Consolidated Overheating Ratio (COR), a metric representing localized thermal excess. SHAP values guide the model to prioritize impactful adjustments that reduce COR while improving interpretability.

To ensure practicality, the method integrates nearest neighbor constraints during counterfactual generation, yielding parameter suggestions that are realistic and data grounded. We introduce SHAP and Neighbor-Enhanced Counterfactual Explanations (SHANCE), which addresses the limitation of conventional CE methods that may propose infeasible recommendations.

A key innovation lies in the addition of a physics-based constraint: volume energy density (VED). This constraint ensures that any suggested power and speed combination remains within manufacturable bounds, aligning with material and process requirements. It acts as a filter to exclude counterfactuals that could result in under- or over-exposure, bridging the gap between data-driven insights and physical feasibility.

The framework was validated using both MNIST and real-world AM datasets. On MNIST, SHANCE improved the realism and diversity of counterfactuals. Applied to AM data, it successfully reduced COR while maintaining process constraints, demonstrating its potential for practical parameter optimization.

By combining interpretability, data proximity, and physical validation, this framework offers a robust and explainable tool for mitigating overheating in AM and improving manufacturing outcomes.

3 - Enabling Comparability: a Benchmarking Framework for Reinforcement Learning in Manufacturing Systems

Lingxiang Yun, University of North Carolina at Charlotte, Charlotte, NC, United States

Reinforcement Learning (RL) has emerged as a promising approach for real-time decision-making in manufacturing systems under uncertainty. However, the lack of standardized simulation environments has hindered the comparability and reproducibility of RL-based methods developed by different research groups. This project introduces a benchmarking framework designed to address this gap by providing a modular, Python-based simulation environment tailored for manufacturing system control tasks. The framework is compatible with popular RL libraries such as Stable-Baselines3 and supports integration with custom algorithms built in PyTorch or TensorFlow. By offering a shared platform for experimentation, this benchmark aims to accelerate progress in RL research for manufacturing and promote fair, transparent evaluation across studies.

WB66

Building B Level 4 B406

Forecasting and Anomaly Detection

Contributed Session

Contributed

Chair: Seungye Lee, Ewha WOMANS UNIVERSITY, Seoul, N/A, Korea, Republic of

1 - An Empirical Investigation of Efficient Anomaly Detection Techniques for Resource-Constrained Systems

Hyungkwon Lee, KAIST, Daejeon, Korea, Republic of, Hyewon Ryu, Youngho Kim, Hangeong Park, Jinuk Oh, Jong-Seok Lee

As anomaly detection has been actively studied, industrial and defense applications increasingly seek to implement anomaly detection systems. However, these systems are typically well-maintained, making anomalies extremely rare or nonexistent in operational data. This scarcity complicates training supervised machine learning models, thus unsupervised methods trained solely on normal data often remain the only practical solution. Additionally, strict hardware constraints in real-world settings—such as security-sensitive systems, thermally restricted equipment, and low-performance computing devices—often make modern deep learning methods with high computational demands impractical. This research aims to support practitioners by examining anomaly detection methods characterized by low computational complexity and quantitatively evaluating their efficiency. We analyze several lightweight algorithms across diverse anomaly scenarios, utilizing both synthetic and limited real anomaly datasets. Our study compares detection performance and computational resource usage under constrained hardware conditions. We also empirically investigate whether the frequently mentioned trade-off between detection accuracy and computational efficiency truly exists. The findings of this study provide practical insights for practitioners designing effective anomaly monitoring systems within resource-limited environments.

2 - An Uncertainty-Aware Physics-Integrated Mixture of Experts for Melt Pool Prediction in Metal Additive Manufacturing

Kyudong Kim, University at Buffalo, Buffalo, NY, United States, Yuxin Yang, Yu Jin

Accurate prediction of melt pool dynamics is essential for ensuring product quality and process reliability in metal additive manufacturing (AM). While laser scan strategies (i.e., laser trajectory) and process parameters significantly influence melt pool behavior, many existing models did not fully capture the relationship between the scan strategy and quality. To address this gap, we propose an uncertainty-aware, physics-integrated Mixture of Experts (UP-MoE) framework tailored for melt pool size prediction for a laser power bed fusion process. The architecture comprises multiple specialized experts, dedicated to capturing distinct aspects of the underlying physics: spatio-temporal melt pool evolution, energy density, radiation pressure, and power distribution. These experts collectively learn the complex interactions between process parameters and melt pool size. To enable principled uncertainty quantification, a Gaussian Process (GP) is integrated with the MoE output, offering reliable confidence estimates critical for process monitoring and control. Numerical studies demonstrate the accuracy and precision of the proposed approach by explicitly modeling scan strategy variations and embedding domain-informed physics into the learning process.

3 - A Comparative Study of Fractional Grey Models for Time Series Forecasting

Gazi Duman, University of New Haven, West Haven, CT, United States, Elif Kongar

Fractional Grey Models represent a significant advancement in Grey System Theory, offering improved flexibility and forecasting accuracy in environments characterized by limited and uncertain data. This study conducts a systematic comparative analysis of several fractional grey models, employing distinct definitions of fractional derivatives, including the Caputo, Riemann-Liouville, Hausdorff and Conformable approaches. The models are evaluated using a benchmark time series dataset, with performance assessed across various criteria such as prediction accuracy, robustness, and computational efficiency. The findings contribute to a deeper understanding of the theoretical and practical implications of fractional calculus in grey modeling and offer evidence-based recommendations for model selection in various applications.

4 - Identifying Key Attributes Influencing Short-Term Rental Occupancy Rates: Case of Airbnb

Ehsan Ahmadi, Mercer University, Atlanta, GA, United States, Sanelisiwe Amanda Nkomo, Reza Maihami

Optimizing the profitability of listings on platforms such as Airbnb is essential in the evolving housing and rental market. Occupancy rates can change due to various factors, such as price, location, host practices, property features, and guest engagement. This study explores the key attributes of higher occupancy rates by analyzing Airbnb listing data. Utilizing data mining and machine learning techniques, the study explores patterns and insights that reveal the most significant factors influencing short-term rental occupancy rates. High occupancy rates, an average of 58% are strongly associated with guest engagement efforts such as the number of reviews, average low prices of \$180, and limiting listing capacity to at least 3 guests. These insights help as a guide for short-term rental hosts in improving occupancy rates and property performance through strategic decisions.

5 - A Patent Location Model for Identifying New Technology Opportunities

Jinseob Kim, Seoul National University, Seoul, Korea, Republic of, Sungjoo Lee

The growing emphasis on evidence-based R&D strategy necessitates determining which patents constitute the core of a technology domain, which actors possess the most extensive and valuable patent portfolios, and where latent R&D opportunities remain. The existing approaches rely heavily on citation, co-classification, or text information; they rarely consider the location of patents within a patent landscape considering technology coverage of patented inventions, which is essential to evaluate the value of patents. This study closes that gap by embedding patent corpora in a high-dimensional semantic space and fulfills the research objective with a facility-location problem. First, a p-center formulation selects a set of patents that minimizes the maximum distance to all others, thereby operationalizing core-technology detection with respect to technological content coverage. Here, the core technology is defined as a patent that can best represent the technological content within technological fields. Next, aggregating the resulting coverage regions by assignee and its country yields quantitative metrics of the actor's technological influence, weighted by forward citations, family size, and claim breadth to measure the technological values. Finally, to identify white-space opportunities, we solve a continuous k-median problem whose coverage gaps highlight candidate coordinates for future invention; each coordinate is reconverted to text via large language model prompting and nearest-neighbor patent sampling, turning abstract vectors into interpretable concept sketches. The framework integrates the facility-location concept with modern language AI to deliver reproducible, content-aware analytics, enabling a novel approach to strategic innovation planning.

6 - Improving Few-Shot Tabular Regression via Latent Task Discovery and Meta-Learning

Yujung Byun, KAIST, Daejeon, Korea, Republic of, Jong-Seok Lee

In many real-world tabular regression problems, data points may actually come from multiple latent tasks or domains, even if we initially treat the dataset as a single task. This hidden heterogeneity can degrade a model's performance and hamper generalization, especially when adapting to new tasks with only a few samples. If the task or domain labels were known, one could apply meta-learning techniques to train models that adapt quickly to each task. However, in our scenario these task structures are not observed, making it challenging to directly use meta-learning. To overcome this challenge, we propose a two-stage framework that discovers latent task structures from data and then leverages meta-learning for improved few-shot regression. In the first stage, our method clusters the training samples into pseudo-tasks based on their prediction error patterns, without using any ground-truth task labels. Samples with similar error behavior are grouped together, as they likely share an underlying function or domain. In the second stage, each pseudo-task is treated as a separate task for meta-learning, training the model to quickly adapt its predictions to each discovered task structure. Through this two-stage process, the model effectively captures task-specific patterns in the data and generalizes better to unseen data or new tasks with only a few examples for adaptation.

7 - Solar Flare Detection Based on Multiple Wavelengths Data

Seungye Lee, Ewha Womans University, Seoul, Korea, Republic of, Dongil Kim

Solar phenomena are observed once every few seconds and stored as a large amount of data. The stored data is used for space environment forecasts. Solar flares, which are solar phenomena that emit strong energy and bright light, can cause fatal damage to social network systems. So accurate detection is necessary. Most of the existing studies use single-form data and basic deep learning models, so it is essential to introduce a new method for improving automatic detection. In this study, we trained a multimodal flare-detection model using solar full-disk images acquired at multiple wavelengths. The proposed method is expected to enable efficient data accumulation with high accuracy by learning solar flare observation data from various aspects.

WB67

Building B Level 4 B407

Robust Operations and Management in Energy Systems

Invited Session

Data Mining

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

Co-Chair: Ramteen Sioshansi, Carnegie Mellon University, Pittsburgh, PA, United States

Co-Chair: Shuyi Chen, Carnegie Mellon University, Pittsburgh, 15213, United States

1 - Dynamic Operating Envelopes for Flexible Management of Distributed Energy Resources

Zhaoyu Wang, Iowa State University, Ames, IA, United States

The increasing penetration of distributed energy resources (DERs) brings both flexibility and uncertainty to power distribution system operations. Under traditional firm interconnection rules, DERs are either allowed full access to the grid or denied interconnection altogether, based on the hosting capacity. Conventional hosting capacity ignores time-varying operational conditions of distribution networks, leading to underutilization of DERs and grid infrastructure. Dynamic operating envelope (DOE) allows the flexible interconnection of DERs. Utilities can dynamically adjust DER power export/import power limits according to real-time grid conditions, such as load fluctuation, voltage level, etc. This talk will introduce a mathematical formulation of DOE, which is widely implemented in commercial advanced distribution management systems of major software vendors. Solving such problems for large-scale distribution grids is extremely challenging due to the non-convexity,

computational burden, and the online operation requirements. In this talk, we will discuss and compare different solution algorithms such as gradient-based, linear programming, and AI-based methods.

2 - Attention is All You Need to Optimize Wind Farm Operations and Maintenance

Iman Kazemian, Wayne state university, Detroit, MI, United States, Murat Yildirim, Paritosh Ramanan

Operations and maintenance (O&M) is a fundamental problem in wind energy systems with far reaching implications for reliability and profitability. Optimizing O&M is a multi-faceted decision optimization problem that requires a careful balancing act across turbine level failure risks, operational revenues, and maintenance crew logistics. The resulting O&M problems are typically solved using large-scale mixed integer programming (MIP) models, which yield computationally challenging problems that require either long-solution times, or heuristics to reach a solution. To address this problem, we introduce a novel decision-making framework for wind farm O&M that builds on a multi-head attention (MHA) models, an emerging artificial intelligence method that are specifically designed to learn in rich and complex problem settings. The development of proposed MHA framework incorporates a number of modeling innovations that allows explicit embedding of MIP models within an MHA structure. The proposed MHA model (i) significantly reduces the solution time from hours to seconds, (ii) guarantees feasibility of the proposed solutions considering complex constraints that are omnipresent in wind farm O&M, (iii) results in significant solution quality compared to the conventional MIP formulations, and (iv) exhibits significant transfer learning capability across different problem settings.

3 - Distributionally Robust Planning of Hydrogen-Electrical Microgrids for Sea Islands

Yuchen Dong, Xi'an Jiaotong University, Xi'an, China, People's Republic of, Zhengwen He, Xiaoyu Cao, Bo Zeng

This paper presents a distributionally robust planning method for hydrogen-electrical microgrids over islands, where the cross-island energy exchange is supported by a maritime hydrogen transport network. This planning problem is complicated due to heterogeneous off-shore wind-driven uncertainties (i.e., renewable power, transport availability, demand fluctuations, and grid faulting), a subset of which exhibit endogenous uncertainty, as they can be affected by proactive measures (e.g., grid hardening) or infrastructure investment. To capture these features, a two-stage distributionally robust optimization (DRO) model is developed considering decision-dependent uncertainty (DDU), which encompasses variation of the underlying distributional ambiguity due to the change of the first stage decisions. Notably, the complete recourse property is missing, which is often neglected in existing DRO studies. Nevertheless, different from the case for land-based microgrids, this issue is critical and fundamental for sea island systems due to their particular physical and logistical requirements. To address these issues, we develop a C&CG algorithm that is customized with strong cutting planes to handle DRO with a varying DDU ambiguity set and feasibility requirements. Numerical results demonstrate the cost-effectiveness and resilience of the proposed planning framework, along with the nontrivial improvements of the algorithm in both solution accuracy and computational efficiency.

4 - Adaptive Robust Optimization with Data-Driven Uncertainty for Enhancing Distribution System Resilience

Shuyi Chen, Carnegie Mellon University, Pittsburgh, PA, United States, Shixiang Zhu, Ramteen Sioshansi

Extreme weather events are placing growing strain on electric power systems, exposing the limitations of purely reactive responses and prompting the need for proactive resilience planning. However, existing approaches often rely on simplified uncertainty models and decouple proactive and reactive decisions, overlooking their critical interdependence. This paper proposes a novel tri-level optimization framework that integrates proactive infrastructure investment, adversarial modeling of spatio-temporal disruptions, and adaptive reactive response. We construct high-probability, distribution-free uncertainty sets using conformal prediction to capture complex and data-scarce outage patterns. To solve the resulting nested decision problem, we derive a bi-level reformulation via strong duality and develop a scalable Benders decomposition algorithm. Experiments on both real and synthetic data demonstrate that our approach consistently outperforms conventional robust and two-stage methods, achieving lower worst-case losses and more efficient resource allocation, especially under tight operational constraints and large-scale uncertainty.

WB68

Building B Level 4 B408

Human-centered Decision-making in Modern Service Operations

Invited Session

Data Mining

Chair: Wei Gu, Carnegie Mellon University, Pittsburgh, 15217, United States

1 - Epidemic Time Series Forecasting in the Era of Machine Learning

Wei Jin, Emory University, Atlanta, GA, United States

As machine learning continues to revolutionize various scientific domains, its impact on epidemic time series forecasting has become increasingly significant. This talk explores the intersection of advanced machine learning techniques with the pressing needs of epidemic forecasting. We will address three key challenges in this field: computational efficiency, data scarcity, and the accurate modeling of spatial-temporal dynamics. First, I will present a solution to enable lighter, faster models to perform at the level of more complex models, crucial for timely epidemic responses. Next, I will discuss the utilization of extensive pre-training on epidemic data to improve the model's predictive accuracy and generalization across varied epidemic scenarios. Finally, I will introduce a graph neural ODE approach for accurately modeling the continuous and dynamic spread of diseases across regions, and outline future research directions in this evolving field of machine learning for epidemic forecasting.

2 - Dynamic Delegation to Achieve Human-AI Complementarity

Wei Gu, Carnegie Mellon University, Pittsburgh, PA, United States, Shixiang Zhu, Michael Lingzhi Li

Humans bring judgment and oversight, while AI offers scale and efficiency. To utilize the complementary strength of both human and AI, we develop a dynamic delegation strategy for adaptive agents that intelligently manage human-AI collaboration. The proposed method optimizes when to act autonomously, defer to humans, or involve human oversight based on the tradeoff between system accuracy and cost. The interaction between human and AI is modeled as a generalized Nash equilibrium and reformulated as a quasi-variational inequality to capture human behavioral adaptation. We provide the conditions under which human and AI can cooperate, and show when the proposed dynamic delegation strategy will outperform static delegation such as fully automated decision-making or mandatory human review. The effectiveness of the proposed method is validated in numerical experiments. Our results offer insights on balancing automation with human input toward building responsible, efficient, and deployable agentic AI systems.

3 - A Semidefinite Relaxation for Copositive Dual Pricing in Discrete Energy Markets

Lauren Henderson, Clemson University, Clemson, SC, United States, Cheng Guo, Ryan Cory-Wright, Boshi Yang

Electricity markets often rely on solving unit commitment (UC) problems to determine generator dispatch and pricing, but non-convexities in these problems lead to challenges in deriving efficient prices. We propose a new pricing scheme, derived from relaxing the completely positive program (CPP) formulation of the UC problem to a semidefinite program (SDP). We show that our SDP is at least as strong as an alternative UC SDP formulation in the literature (Fattahi et al., 2017). We introduce coordination agents that model cross-terms in the semidefinite relaxation, which gives our pricing scheme economically desirable properties like revenue adequacy and cost recovery. Additionally, we provide perturbation bounds using the envelope theorem, which describes the optimal UC value as a function of the cost of starting up or shut down a generator. We perform numerical experiments on a variety of IEEE UC instances to demonstrate the efficacy of our pricing scheme. Specifically, we show that our SDP relaxation yields a much tighter approximation to the UC value function than the linear programming-based relaxation. We also see promising results of lower uplift payments compared with the commonly-used fixed binary pricing scheme.

4 - Operator Learning for Human-Centered Indoor Air Quality Control in Smart Buildings

Yuxin Bian, University of California San Diego, La Jolla, CA, United States, Oliver Schmidt, Yuanyuan Shi

Ensuring healthy indoor air quality while minimizing energy consumption is a central challenge in smart building control. Traditional ventilation strategies often overlook fine-grained airflow dynamics, leading to inefficient or suboptimal performance. While Computational Fluid Dynamics (CFD) provides high-fidelity modeling of airflow, its prohibitive computational cost hinders real-time decision-making. In this work, we propose a data-driven operator learning framework that captures complex indoor airflow patterns and supports intelligent, health-aware ventilation control. Leveraging high-resolution CFD data, we train a neural operator transformer to learn the mapping from control inputs—such as vent angles and supply rates—to full-field airflow and CO₂ distributions. This learned operator enables gradient-based, real-time

optimization of ventilation settings to simultaneously ensure occupant health and reduce energy use. We evaluate our method across various indoor scenarios and compare it with rule-based control, data-driven approaches based on average CO₂ modeling, and reduced-order modeling techniques. Our approach consistently maintains safe air quality levels while achieving substantial energy savings, demonstrating its promise for scalable, human-aware smart building management.

Wednesday, October 29, 11:00 AM - 12:00 PM

Building A Level 4 Auditorium

Plenary - Learning to Optimize: Foundations and Industrial Impact

Plenary Keynote

Plenary

1 - Learning to Optimize: Foundations and Industrial Impact

Pascal Van Hentenryck, ISyE Georgia Tech, Atlanta, GA, United States

In many industry settings, the same optimization problem is solved repeatedly for instances taken from a distribution that can be learned or forecasted. Indeed, such parametric optimization problems are ubiquitous in applications over complex infrastructures such as electrical power grids, supply chains, manufacturing systems, and transportation networks. The scale and complexity of these applications, and increasingly volatile environments they operate on, have grown significantly in recent years, challenging traditional optimization approaches. This talk studies how to fuse machine learning and optimization for speeding up these parametric optimization problems and meeting real-time constraints present in practice. It first reviews the concept of optimization proxies that learn the input/output mappings of parametric optimization problems, computing near-optimal feasible solutions and providing quality guarantees orders of magnitude faster. The talk also presents how to “learn to optimize” highly complex optimization problems, fusing optimization methodologies with supervised learning and reinforcement learning. The impact of the methodologies are highlighted on industrial problems in grid optimization, end-to-end supply chains, logistics, and transportation systems. The talk should reveal beautiful connections between machine learning and optimization, that leverage fundamental theoretical results to push the practice of optimization.